Copulas: An Introduction
Part II: Models

Johan Segers

Université catholique de Louvain (BE)
Institut de statistique, biostatistique et sciences actuarielles

Columbia University, New York City
9–11 Oct 2013

Copulas: An Introduction
Part II: Models

Archimedean copulas

Extreme-value copulas

Elliptical copulas

Vines

Copulas: An Introduction
Part II: Models

Archimedean copulas

Extreme-value copulas

Elliptical copulas

Vines

The (in)famous Archimedean copulas

- By far the most popular (theory & practice) class of copulas
- Plenty of parametric models
  - Gumbel, Clayton, Frank, Joe, Ali–Mikhail–Haq, …
- Building block for more complicated constructions:
  - Nested/Hierarchical Archimedean copulas
  - Vine copulas
  - Archimax copulas
  - …
- Mindless application of (Archimedean) copulas has drawn many criticisms on the copula 'hype'

Laplace transform of a positive random variable

Recall the Laplace transform of a random variable $Z > 0$:

$$\psi(s) = \mathrm{E}[\exp(-sZ)] = \int_0^\infty e^{-sz}\,\mathrm{d}F_Z(z), \qquad s \in [0, \infty]$$

A distribution on $(0, \infty)$ is identified by its Laplace transform.

Ex. Show the following properties:

- $0 \le \psi(s) \le 1$
- $\psi(0) = 1$ and $\psi(\infty) = 0$.
- $(-1)^k \mathrm{d}^k \psi(s) / \mathrm{d}s^k \ge 0$ for all integer $k \ge 1$.
- In particular, ψ is nonincreasing ($k = 1$) and convex ($k = 2$).

Survival functions in proportional hazards model: The Laplace transform of the frailty appears

Independent unit exponential random variables $Y_1, \ldots, Y_d$.

Survival times $X_1, \ldots, X_d$ are affected by a common 'frailty' $Z > 0$:

$$X_j = Y_j / Z$$

Marginal and joint survival functions:

$$\begin{aligned}
\Pr[X_j > x_j] &= \mathrm{E}[e^{-x_j Z}] \\
&= \psi(x_j) \\
\Pr[X_1 > x_1, \ldots, X_d > x_d] &= \mathrm{E}[e^{-(x_1 + \cdots + x_d) Z}] \\
&= \psi(x_1 + \cdots + x_d)
\end{aligned}$$

In proportional hazards models, survival copulas are Archimedean

The survival copula of $\boldsymbol{X}$ is Archimedean with generator ψ:

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

Ex. Show the above formula.

Ex. Show that replacing Z by βZ for a constant $\beta > 0$ changes ψ but does *not* change the copula.

Ex. Pick your favourite (discrete/continuous) distribution on $(0, \infty)$, compute or look up its Laplace transform, and compute the associated Archimedean copula. If it doesn't exist yet, name it after yourself and publish a paper about it.

A Gamma frailty induces the Clayton copula

If $Z \sim \text{Gamma}(1/\theta, 1)$, with $0 < \theta < \infty$, then

$$\psi(s) = \int_0^\infty e^{-sz} \, \frac{z^{1/\theta-1} e^{-z}}{\Gamma(1/\theta)} \, \mathrm{d}z = (1+s)^{-1/\theta}$$

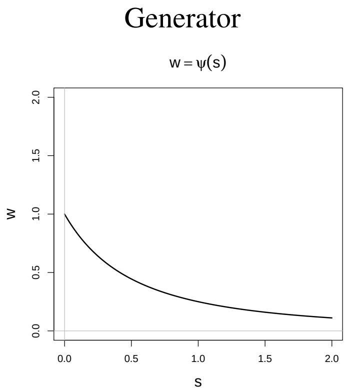
and the resulting survival copula is Clayton:

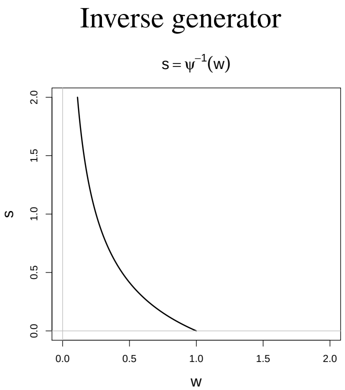
$$\bar{C}(\boldsymbol{u}) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}$$

Ex. Check the above formulas.

Ex. How to use the frailty representation to sample from a Clayton copula?

Generator of the Clayton copula





$$w = \psi(s) = (1+s)^{-1/\theta}$$

$$s = \psi^{-1}(w) = w^{-\theta} - 1$$

Formal definition of an Archimedean copula

A copula C is Archimedean if there exists $\psi : [0, \infty] \to [0, 1]$ such that

$$C(\boldsymbol{u}) = \psi\big(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\big)$$

For C to be a copula, it is sufficient and necessary that ψ satisfies

- $\psi(0) = 1$ and $\psi(\infty) = 0$
- ψ is d-monotone, i.e.
  - $(-1)^k \mathrm{d}^k \psi(s)/\mathrm{d}s^k \geq 0$ for $k \in \{0, \ldots, d-2\}$
  - $(-1)^{d-2} \mathrm{d}^{d-2} \psi(s)/\mathrm{d}s^{d-2}$ is decreasing and convex

Equivalently, there should exists a random variable $Z > 0$ such that

$$\psi(s) = \mathrm{E}\left[\left(1 - \frac{sZ}{d-1}\right)_+^{d-1}\right]$$

i.e. ψ is the Williamson d-transform of the rv $(d-1)/Z$.

Standard examples

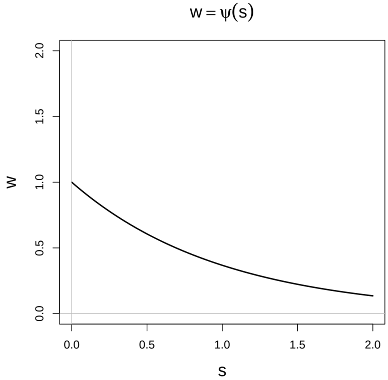
Ex. The independence copula $\Pi(\boldsymbol{u}) = u_1 \cdots u_d$ is Archimedean.

- What is its generator ψ?
- What is the frailty variable Z?

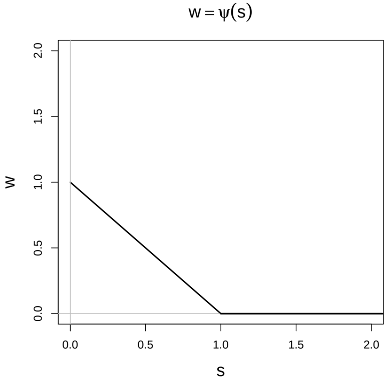
Ex. The Fréchet–Hoeffding lower bound $W(u, v) = \max(u + v - 1, 0)$ is Archimedean too. What is its generator ψ?
[This ψ is *not* a Laplace transform; it is 2-monotone but not d-monotone for $d \geq 3$.]

Ex. One can show that the Fréchet–Hoeffding upper bound $M(\boldsymbol{u}) = \min(u_1, \ldots, u_d)$ is *not* Archimedean. Still, show that the Clayton copula with $\theta \to \infty$ converges to M.

Common generator functions



$$\Pi(\boldsymbol{u}) = u_1 \cdots u_d$$
$$\psi(s) = e^{-s}$$



$$W(u, v) = \max(u + v - 1, 0)$$
$$\psi(s) = \max(1 - s, 0)$$

Bivariate Archimedean copulas as binary operators

A bivariate Archimedean copula induces a binary operator

$$[0,1] \times [0,1] \to [0,1] : (u,v) \mapsto C(u,v)$$

which is commutative and associative:

$$\begin{aligned} C(u,v) &= C(v,u), \\ C(u,C(v,w)) &= C(C(u,v),w) \end{aligned}$$

endowing $[0,1]$ with a semi-group structure.

Link with the theory of associative functions (Abel, Hilbert).

Derived quantities

Conditional cdf:

$$\dot{C}_j(\boldsymbol{u}) = \frac{\psi'\bigl(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\bigr)}{\psi'\bigl(\psi^{-1}(u_j)\bigr)}$$

Pdf, provided ψ is d times continuously differentiable

$$c(\boldsymbol{u}) = \frac{\psi^{(d)}\bigl(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\bigr)}{\prod_{j=1}^{d} \psi'\bigl(\psi^{-1}(u_j)\bigr)}$$

Ex. Show these formulas.

Yet another probability integral transform: Kendall distribution functions

Bivariate cdf H, continuous margins F and G, copula C.

The Kendall distribution of a random pair $(X, Y) \sim H$ is the cdf of the rv

$$W = H(X, Y) = C(F(X), G(Y)) = C(U, V)$$

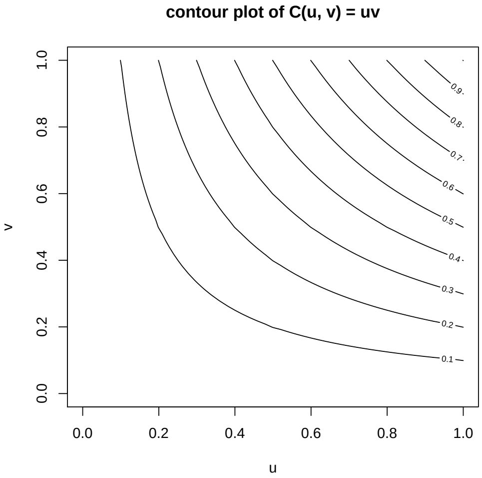
It only depends on H through C:

$$K_C(w) = \Pr(W \le w) = \int_{[0,1]^2} \mathbf{1}\{C(u, v) \le w\} \, \mathrm{d}C(u, v), \qquad w \in [0, 1]$$

It is linked to Kendall's tau via

$$\mathrm{E}[W] = \int_0^1 w \, \mathrm{d}K_C(w) = \int_{[0,1]^2} C(u, v) \, \mathrm{d}C(u, v) = \frac{1 + \tau}{4}$$

Kendall distribution functions: The C-probability below a C-level curve



$$K(w) = \int_{[0,1]^2} \mathbf{1}\{C(u,v) \le w\}\, \mathrm{d}C(u,v)$$

Bivariate Archimedean copulas are identified by their Kendall distribution function

The Kendall distribution function of a bivariate Archimedean copula with inverse generator $\phi = \psi^{-1} : (0, 1] \to [0, \infty)$ is

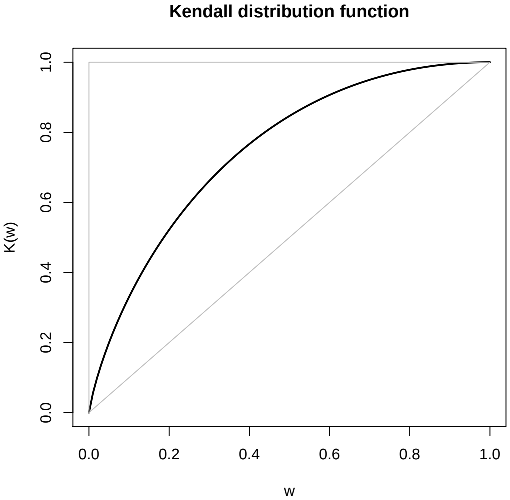
$$K(w) = w - \lambda(w),$$
$$\lambda(w) = \frac{\phi(w)}{\phi'(w)} = \frac{1}{\mathrm{d} \log \phi(w)/\mathrm{d}w} \leq 0$$

Up to a multiplicative constant, ϕ and thus ψ can be reconstructed from λ.

Ex. Show the following properties:

- $K_\Pi(w) = w - w \log(w)$ (independence)
- $K_W(w) = 1$ (Fréchet–Hoeffding lower bound)
- $K_M(w) = w$ (Fréchet–Hoeffding upper bound)
- $w \leq K(w) \leq 1$

Kendall distribution functions: Stochastically smaller than the uniform one



The tail behaviour of a bivariate Archimedean copula can be read off from the inverse generator function

Coefficient of lower tail dependence:

$$\lambda_L(C) = \lim_{w \downarrow 0} \frac{C(w, w)}{w} = 2^{-1/\theta_0},$$

$$\text{where } \theta_0 = - \lim_{w \downarrow 0} \frac{w\, \phi'(w)}{\phi(w)} \in [0, \infty]$$

Coefficient of upper tail dependence:

$$\lambda_U(C) = \lambda_L(\bar{C}) = 2 - 2^{1/\theta_1},$$

$$\text{where } \theta_1 = - \lim_{w \downarrow 0} \frac{w\, \phi'(1 - w)}{\phi(1 - w)} \in [1, \infty]$$

$\Rightarrow$ Construction of models with different upper and lower tails

Archimedean copulas enjoy many symmetries

Let $(U_1, \ldots, U_d) \sim C$ and C is Archimedean with generator ψ.

- Permuation symmetry: For any permuation σ of $\{1, \ldots, d\}$,

$$(U_{\sigma(1)}, \ldots, U_{\sigma(d)}) \stackrel{d}{=} (U_1, \ldots, U_d)$$

- Closure of margins: For any subset $1 \le j_1 < \cdots < j_k \le d$,

$$(U_{j_1}, \ldots, U_{j_k}) \sim k\text{-variate Archimedean, same generator } \psi$$

Symmetry is a blessing (simplicity) and a curse (lack of flexibility).

The only radially symmetric Archimedean copula ($C = \bar{C}$) is the Frank copula.

Escaping from permutation symmetry: Nested Archimedean copulas

$$
\begin{tikzcd}
 & \psi_0 \arrow[dl] \arrow[dr] & & \\
U_1 & & \psi_{23} \arrow[dl] \arrow[dr] & \\
 & U_2 & & U_3
\end{tikzcd}
$$

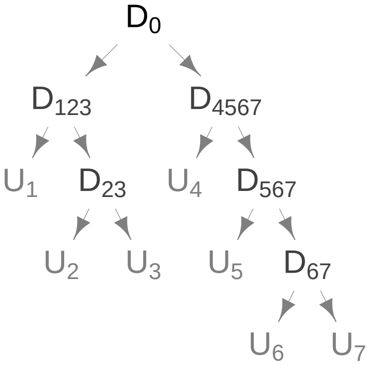
Trivariate copula:

$$
\begin{aligned}
C(u_1, u_2, u_3) &= C_{\psi_0}\big(u_1, C_{\psi_{23}}(u_2, u_3)\big) \\
&= \psi_0\Big(\psi_0^{-1}(u_1) + \psi_0^{-1}\big(\psi_{23}(\psi_{23}^{-1}(u_2) + \psi_{23}^{-1}(u_3))\big)\Big)
\end{aligned}
$$

Bivariate margins:

- (U_1, U_2) Archimedean with generator ψ_0
- (U_1, U_3) Archimedean with generator ψ_0
- (U_2, U_3) Archimedean with generator ψ_{23}

Nested Archimedean copulas: Hierarchical dependence structure



Dependence at deeper levels must be stronger than at higher levels:
Sufficient nesting condition on generator functions

Archimedan copulas: Some literature

Alsina, C., M. J. Frank, and B. Schweizer (2006). *Associative functions: triangular norms and copulas*. New Jersey: World Scientific.

Charpentier, A. and J. Segers (2009). Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis 2009*, 1521–1537.

Genest, C., J. Nešlehová, and J. Ziegel (2011). Inference in multivariate Archimedean copula models. *Test 20*(2), 223–256.

Genest, C. and L.-P. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association 88*(423), 1034–1043.

McNeil, A. J. and J. Nešlehová (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *The Annals of Statistics 37*(5B), 3059–3097.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer. Chapter 4.

Okhrin, O., Y. Okhrin, and W. Schmid (2013). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics 173*(2), 189–204.

Copulas: An Introduction
Part II: Models

How to define the maximum of a multivariate sample?

Consider iid $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from F with continuous margins $F_1, \ldots, F_j$ and copula C.

Vector of component-wise maxima:

$$\begin{aligned} \boldsymbol{M}_n &= (M_{n,1}, \ldots, M_{n,d}) \\ M_{n,j} &= \max(X_{1,j}, \ldots, X_{n,j}), \qquad j \in \{1, \ldots, d\} \end{aligned}$$

In general, $\boldsymbol{M}_n \not\in \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$.

Ex. Draw a scatter plot of a bivariate sample and locate the point representing the pair of maxima.

The copula of the vector of sample maxima

The joint and marginal cdfs of $\boldsymbol{M}_n$:

$$\Pr(\boldsymbol{M}_n \le \boldsymbol{x}) = F^n(\boldsymbol{x}),$$
$$\Pr(M_{n,j} \le x_j) = F_j^n(x_j)$$

The copula of $\boldsymbol{M}_n$:

$$C_n(\boldsymbol{u}) = C(u_1^{1/n}, \ldots, u_d^{1/n})^n$$

Ex. Prove the above equations.

Ex. If $d = 2$ and $X_{i,2} = -X_{i,1}$, we find the Clayton copula with $\theta = -1/n$.

Extreme-value copulas: Limits of copulas of sample maxima

A copula is an extreme-value copula if it can arise in the limit

$$C_\infty(\boldsymbol{u}) = \lim_{n\to\infty} C(u_1^{1/n}, \ldots, u_d^{1/n})^n$$

Extreme-value copulas are max-stable:

$$C_\infty(u_1^{1/k}, \ldots, u_d^{1/k})^k = C_\infty(\boldsymbol{u})$$

Conversely, max-stable copulas are extreme-value copulas.

The only max-stable Archimedean copula is the Gumbel copula

Ex. Show that the *Gumbel* copula is max-stable:

$$C_\theta(\boldsymbol{u}) = \exp[-\{(-\log u_1)^\theta + \cdots + (-\log u_d)^\theta\}^{1/\theta}], \qquad \theta \in [1, \infty]$$

Special cases:

$\theta = 1$ Independence
$\theta = \infty$ Fréchet–Hoeffding upper bound

Maxima versus minima: Just switch to survival copulas

Everything can be repeated for minima, but the formulas get unwieldy

- Apply inclusion/exclusion formulas.

Conceptually, just switch to survival copulas:

$$C_\infty \textit{ is max/min-stable} \iff \bar{C}_\infty \textit{ is min/max-stable}$$

A solution in practice:
If interest is in minima, change signs and work with maxima.

The domain of attraction of an extreme-value copula

The (max-)domain of attraction of an extreme-value copula C_∞ is the collection of all copulas C such that

$$\lim_{n\to\infty} C(u_1^{1/n}, \ldots, u_d^{1/n})^n = C_\infty(\boldsymbol{u}) \tag{DA}$$

Clearly, $C_\infty \in \mathrm{DA}(C_\infty)$.

Alternative condition for (DA) in terms of behaviour of C near $(1, \ldots, 1)$:

$$\begin{aligned}\lim_{s\downarrow 0} s^{-1}\{1 - C(1 - sx_1, \ldots, 1 - sx_d)\} \\ = \log C_\infty(e^{-x_1}, \ldots, e^{-x_d}) =: \ell(\boldsymbol{x}), \qquad \boldsymbol{x} \in [0, \infty)^d\end{aligned}$$

The limit is called the stable tail dependence function.
[Proof: In (DA), take logarithms and set $s = 1/n$ and $u_j = e^{-x_j}$.]

Archimedean copulas: Attracted by the Gumbel copula

If C is Archimedean with inverse generator $\phi = \psi^{-1}$ and if

$$\exists \lim_{w \downarrow 0} - \frac{w\,\phi'(1-w)}{\phi(1-w)} = \theta_1 \in [1, \infty]$$

then $C \in \mathrm{DA}\big(\text{Gumbel copula } C_{\theta_1}\big)$.

Ex. Show that the *Joe* copula with inverse generator

$$\phi_\theta(w) = -\log\big(1 - (1-w)^\theta\big), \qquad \theta \in [1, \infty),$$

is attracted by the Gumbel copula with parameter θ.

Archimedean survival copulas: Attracted by the Galambos copula

If C is Archimedean with inverse generator $\phi = \psi^{-1}$ and if

$$\exists \lim_{w \downarrow 0} - \frac{w\, \phi'(w)}{\phi(w)} = \theta_0 \in [0, \infty]$$

then $\bar{C} \in \mathrm{DA}(\textit{Galambos} \text{ copula } C_{\theta_0})$, with stdf

$$\ell_\theta(\boldsymbol{x}) = x_1 + \cdots + x_d - \sum_{I \subset \{1,\ldots,d\}, |I| \geq 2} (-1)^{|I|} \left(\sum\nolimits_{j \in J} x_j^{-\theta} \right)^{-1/\theta}$$

Ex. Show that the survival *Clayton* copula with inverse generator

$$\phi_\theta(w) = \frac{w^{-\theta} - 1}{\theta}, \qquad \theta \in [0, \infty),$$

is attracted by the Galambos copula with the same parameter.

Pickands dependence functions: A kind of generator function on the unit simplex

If C_∞ is max-stable, the function A on

$$\Delta_{d-1} = \{\boldsymbol{t} \in [0,1]^d : t_1 + \cdots + t_d = 1\}$$

defined by

$$A(\boldsymbol{t}) = \frac{\log C_\infty(w^{t_1}, \ldots, w^{t_d})}{\log w}$$

does *not* depend on $w \in (0,1)$. We find the Pickands representation

$$C_\infty(w^{t_1}, \ldots, w^{t_d}) = w^{A(\boldsymbol{t})}$$

For bivariate extreme-value copulas, Pickands functions are simple objects

In the bivariate case, identifying $(1-t,t) \equiv t$ and writing

$$(u,v) = (w^{1-t}, w^{t}) \text{ with } w = uv \text{ and } t = \frac{\log(v)}{\log(uv)}$$

we obtain the representation

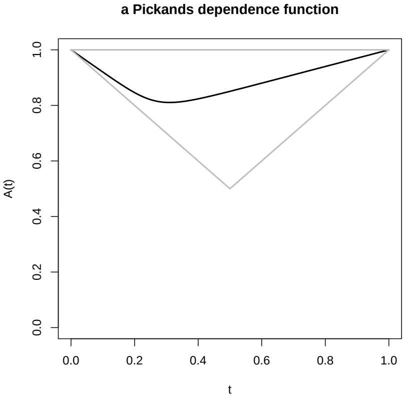
$$C_\infty(u,v) = (uv)^{A(t)}$$

Necessary and sufficient condition on A for C_∞ to be a copula:

- $\max(t, 1-t) \leq A(t) \leq 1$
- A is convex

Ex. Show that if C_∞ as defined above is a copula, it is max-stable.

Bounds for extreme-value copulas



Between independence and complete dependence:

$$\begin{array}{ccccc} uv & \leq & C(u,v) & \leq & \min(u,v) \\ 1 & \geq & A(t) & \leq & \max(t, 1-t) \end{array}$$

The upper and lower bounds are extreme-value copulas too.

Extreme-value copulas: An abundance of parametric models

Ex. Look up the forms of the following extreme-value copulas and visualize their Pickands dependence functions:

- Gumbel aka logistic, and asymmetric extensions
- Galambos aka negative logistic, and asymmetric extensions
- Marshall–Olkin
- Hüsler–Reiss
- t-EV
- Schlather
- ...

Extreme-value copulas: Flexible models for positively associated variables

- Kendall's tau:

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} \, \mathrm{d}A'(t) > 0 \text{ unless independence}$$

- Coefficient of upper tail dependence:

$$\lambda_U = 2\bigl(1 - A(1/2)\bigr) > 0 \text{ unless independence}$$

- Not necessarily symmetric
- Higher dimensions: hierarchical structures possible
- Margins of extreme-value copulas are also extreme-value copulas

Extreme-value copulas: Some literature I

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley. Chapters 8 and 9.

Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics 39*(4), 1963–2006.

Fougères, A., J. Nolan, and H. Rootzén (2009). Models for dependent extremes using stable mixtures. *Scandinavian Journal of Statistics 36*, 42–59.

Genest, C., I. Kojadinovic, J. Nešlehová, and J. Yan (2011). A goodness-of-fit test for extreme-value copulas. *Bernoulli 17*, 253–275.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics 37*(5B), 2990–3022.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In W. H. P. Jaworski, F. Durante and T. Rychlik (Eds.), *Proceedings of the Workshop on Copula Theory and its Applications*, Berlin, pp. 127–146. Springer.

Gudendorf, G. and J. Segers (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference 142*, 373–385.

Extreme-value copulas: Some literature II

Guillotte, S. and F. Perron (2008). A Bayesian estimator for the dependence function of a bivariate extreme-value distribution. *The Canadian Journal of Statistics 36*(3), 383–396.

Kojadinovic, I., J. Segers, and J. Yan (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *The Canadian Journal of Statistics 39*, 703–720.

Peng, L., L. Qian, and J. Yang (2013). Weighted estimation of the dependence function for an extreme-value distribution. *Bernoulli 19*(2), 492–520.

Zhang, D., M. T. Wells, and L. Peng (2008). Nonparametric estimation of the dependence function for a multivariate etreme value distribution. *Journal of Multivariate Analysis 99*(4), 577–588.

Copulas: An Introduction
Part II: Models

Archimedean copulas

Extreme-value copulas

Elliptical copulas

Vines

Elliptical random vectors: Affine transformations of spherically symmetric ones

A random vector $\boldsymbol{X}$ has an elliptical distribution if it can be written

$$\boldsymbol{X} = \mu + \varrho A \boldsymbol{V}$$

- $\mu \in \mathbb{R}^d$
- $\varrho \geq 0$ random
- $A \in \mathbb{R}^{d \times d}$
- $\boldsymbol{V}$ is uniformly distributed on $\{\boldsymbol{v} \in \mathbb{R}^d : v_1^2 + \cdots + v_d^2 = 1\}$
- ϱ and $\boldsymbol{V}$ are independent

Elliptical distributions: Elliptically contoured densities

If

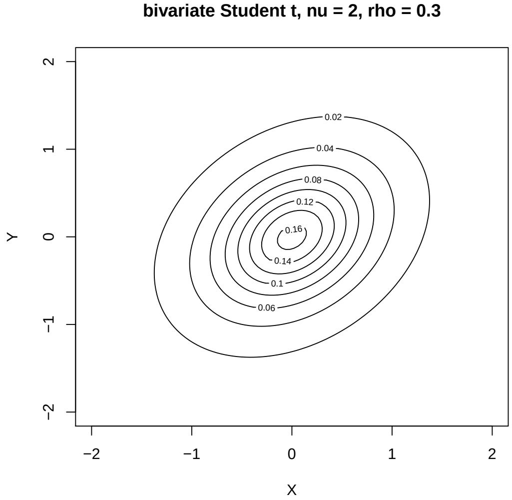
- ϱ has a density f_ϱ
- $\Sigma = AA^\top$ is invertible

then $\boldsymbol{X}$ has a density $f_{\boldsymbol{X}}$ too, and $f_{\boldsymbol{X}}(\boldsymbol{x})$ depends on

- f_ϱ (radial density)
- $\sqrt{(\boldsymbol{x} - \boldsymbol{u})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{u})}$ (Mahalanobis distance)

Contour sets of $f_{\boldsymbol{X}}$ are elliptical.

Densities with elliptical contour lines



Most common elliptical distributions: Gaussian and Student

ϱ	$\boldsymbol{X}$
χ^2_d	Gaussian
$F_{d,\nu}$	Student

Link between both: If

- $\boldsymbol{Z} \sim N_d(\mathbf{0}, \Sigma)$
- $V \sim \chi^2_\nu$
- $\boldsymbol{Z}$ and V are independent

Then $\boldsymbol{X} = \boldsymbol{Z}/\sqrt{V/\nu}$ is Student(0, Σ, ν).

If $\nu \to \infty$, then 'Student' tends to 'Gaussian'.

Meta-elliptical copulas: Copulas of elliptical distributions

A copula is meta-elliptical if it is the copula of an elliptical distribution.

A meta-elliptical copula is itself not an elliptical distribution.
Hence 'meta'; suppressed in practice.

Without loss of generality, we can assume that

- $\mu = 0$
- Σ is a correlation matrix, notation R

Ex. Why?

The Gaussian and Student copulas

Gaussian copula: copula of $\boldsymbol{Z} \sim N_d(0, R)$,

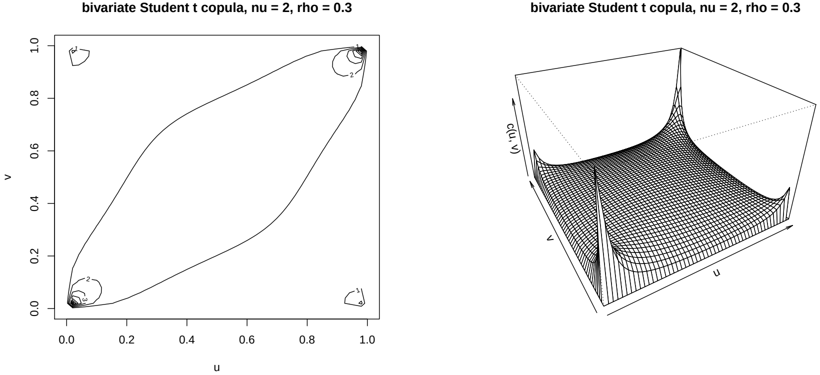
$$C_R^{\text{Gauss}}(\boldsymbol{u}) = \Pr[\Phi(Z_1) \le u_1, \ldots, \Phi(Z_d) \le u_d]$$

Student copula: copula of $\boldsymbol{T} \sim \text{Student}_d(0, R, \nu)$,

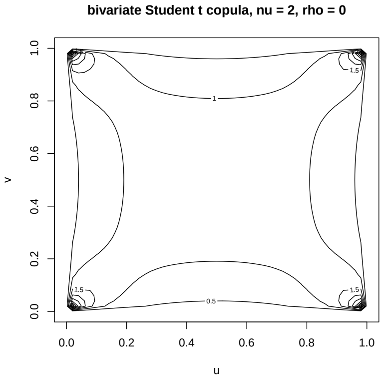
$$C_{R,\nu}^{\text{Student}}(\boldsymbol{u}) = \Pr[t_\nu(T_1) \le u_1, \ldots, t_\nu(T_d) \le u_d]$$

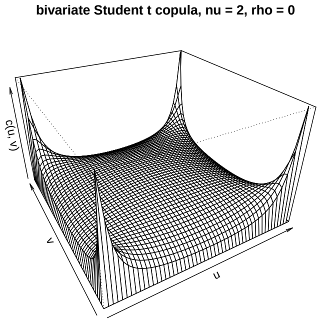
with t_ν the univariate standard Student(ν) cdf.

Elliptical copula densities: Contour lines are not elliptical



Zero correlation implies independence for Gaussian copulas only





Elliptical copulas are convenient to work with

- Densities are explicitly available.
- Pairwise distributions determine the full distribution.
- Lower-dimensional margins are elliptical copulas again.
- If $\boldsymbol{U} \sim C$ is elliptical, then, whatever the radial distribution,

$$\tau(U_j, U_k) = \frac{\arcsin(r_{jk})}{\pi/2} \qquad \text{(Kendall's tau)}$$

- Tail dependence follows from power-law tail of ϱ, e.g.
  - Gaussian copula: asymptotic independence
  - Student copula: $\lambda_L = \lambda_U = 2\, t_{\nu+1}(-\sqrt{(\nu+1)(1-\rho)/(1+\rho)})$

Putting structure on the correlation matrix allows for interpretable models

Factor models: for $\Gamma^{k \times d}$ with $k < d$,

$$\Sigma = \Gamma' \Gamma + \sigma^2 I_d$$

Graphical models: Gaussian with *sparse* inverse matrix R^{-1}

$$(R^{-1})_{jk} = \text{partial correlation of } Z_j \text{ and } Z_k \text{ given the other variables}$$

$\Rightarrow$ Conditional independence graphs.

Elliptical copulas: Some literature

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Fang, H.-B., K.-T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*, 1–16.

Genest, C., A.-C. Favre, J. Béliveau, and C. Jacques (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research 43*, W09401.

Hult, H. and F. Lindskog (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Probab. 34*, 587–608.

Klüppelberg, C. and G. Kuhn (2009). Copula structure analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(3), 737–753.

Copulas: An Introduction
Part II: Models

Archimedean copulas

Extreme-value copulas

Elliptical copulas

Vines

The simplifying assumption: The copula of a conditional distribution

For random variables (X, Y) and a random vector $\boldsymbol{Z}$, assume:

The copula of $(X, Y) \mid \boldsymbol{Z} = \boldsymbol{z}$ *does not depend on* $\boldsymbol{z}$.

Equivalently, assume:

$(F_{X|\boldsymbol{Z}}(X \mid \boldsymbol{Z}), F_{Y|\boldsymbol{Z}}(Y \mid \boldsymbol{Z}))$ *is independent of* $\boldsymbol{Z}$.

- True if $(X, Y, \boldsymbol{Z})$ are jointly Gaussian
  - $(X, Y) \mid \boldsymbol{Z} = \boldsymbol{z}$ is bivariate Gaussian
  - Conditional correlation is partial correlation $\rho_{XY \cdot \boldsymbol{Z}}$, whatever $\boldsymbol{z}$
- Simplifying assumption not verified in general

From the simplifying assumption to vine copulas

Vine copulas or pair-copula constructions:
Combine $d(d-1)/2$ *arbitrary* bivariate copulas into a d-variate copula.

- The bivariate copulas are *not* the bivariate margins.
- They rather arise through repeated conditioning.
- Construction made possible by the simplifying assumption.

From bivariate to conditional densities

Random pair (X, Y):

$$
\begin{aligned}
f_{XY}(x,y) &= c\big(F_X(x), F_Y(y)\big)\, f_X(x)\, f_Y(y) && \text{bivariate} \\
f_{X|Y}(x,y) &= c\big(F_X(x), F_Y(y)\big)\, f_X(x) && \text{conditional}
\end{aligned}
$$

Similarly, but now conditionally on a random vector $\boldsymbol{Z}$:

$$
\begin{aligned}
f_{XY|\boldsymbol{Z}}(x,y \mid z) &= c_{XY|\boldsymbol{Z}}\big(F_{X|\boldsymbol{Z}}(x \mid z), F_{Y|\boldsymbol{Z}}(y \mid z)\big)\, f_{X|\boldsymbol{Z}}(x \mid z)\, f_{Y|\boldsymbol{Z}}(y \mid z) \\
f_{X|Y,\boldsymbol{Z}}(x \mid y,z) &= c_{XY|\boldsymbol{Z}}\big(F_{X|\boldsymbol{Z}}(x \mid z),\ F_{Y|\boldsymbol{Z}}(y \mid z)\big)\, f_{X|\boldsymbol{Z}}(x \mid z)
\end{aligned}
$$

Vines: use this formula iteratively to factorize a multivariate pdf

Ex. Where exactly was the simplifying assumption used?

Vines in dimension three

Taking X_3 as 'pivot' variable:

$$
\begin{aligned}
& f(x_1, x_2, x_3) \\
= \; & f_3(x_3) \\
& f_{2|3}(x_2|x_3) \\
& f_{1|23}(x_1|x_2, x_3) \\
= \; & f_3(x_3) \\
& c_{23}\big(F_2(x_2), F_3(x_2)\big)\, f_2(x_2) \\
& c_{12|3}\big(F_{1|3}(x_1|x_3),\ F_{2|3}(x_2|x_3)\big) \qquad \underbrace{f_{1|3}(x_1|x_3)}_{=c_{13}(F_1(x_1), F_3(x_3))\, f_1(x_1)} \\
= \; & f_1(x_1)\, f_2(x_2)\, f_3(x_3) \\
& c_{13}\big(F_1(x_1),\ F_3(x_3)\big)\, c_{23}\big(F_2(x_2),\ F_3(x_2)\big)\, c_{12|3}\big(\underbrace{F_{1|3}(x_1|x_3)}_{=?},\ F_{2|3}(x_2|x_3)\big)
\end{aligned}
$$

The conditional cdf's follow from the pair copulas too

Conditional cdf:

$$\begin{aligned}
F_{1|3}(x_1|x_3) &= \int_{-\infty}^{x_1} f_{1|3}(x_1'|x_3)\,\mathrm{d}x_1' \\
&= \int_{-\infty}^{x_1} c_{13}\big(\underbrace{F_1(x_1')}_{=u_1},\, F_3(x_3)\big)\, f_1(x_1')\,\mathrm{d}x_1' \\
&= \int_0^{F_1(x_1)} c_{13}\big(u_1, F_3(x_3)\big)\,\mathrm{d}u_1 \\
&= \frac{\partial}{\partial u_3} C_{13}\big(F_1(x_1), u_3\big)\,\bigg|_{u_3=F_3(x_3)}
\end{aligned}$$

Depends on C_{13}, F_1 and F_3

Vines in dimension four

Single out one variable:

$$f(x_1, x_2, x_3, x_4) = \underbrace{f_{234}(x_2, x_3, x_4)}_{\text{trivariate}} \, f_{1|234}(x_1 \mid x_2, x_3, x_4)$$

Decompose the conditional density:

$$f_{1|234}(x_1 \mid x_2, x_3, x_4) = c_{12|34}\big(F_{1|34}(x_1 \mid x_3, x_4), \, F_{2|34}(x_2 \mid x_3, x_4)\big) \\ f_{1|34}(x_1 \mid x_2, x_3)$$

The conditional density $f_{1|34}(x_1 \mid x_3, x_4)$ was treated above.
By the same argument as on the previous slide, the conditional cdf is

$$F_{1|34}(x_1 \mid x_3, x_4) = \frac{\partial}{\partial u_3} C_{13|4}\big(F_{1|4}(x_1 \mid x_4), u_3\big)\Big|_{u_3 = F_{3|4}(x_3 \mid x_4)}$$

In dimension four, six pair copulas are needed

Collecting everything, we find a decomposition in terms of six pair copulas:

Canonical (C) vine $\begin{cases} c_{14}, c_{24}, c_{34} & \text{'ground level'} \\ c_{13|4}, c_{23|4} & \text{'level 1'} \\ c_{12|34} & \text{'level 2'} \end{cases}$

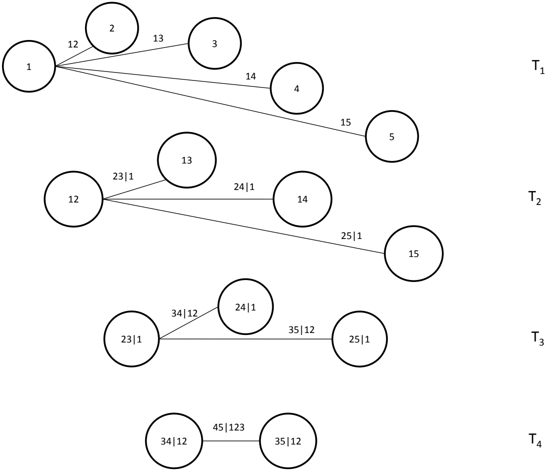
With other choices of the conditioning variables, we would have obtained:

Drawable (D) vine $\begin{cases} c_{12}, c_{23}, c_{34} & \text{'ground level'} \\ c_{13|2}, c_{24|3} & \text{'level 1'} \\ c_{14|23} & \text{'level 2'} \end{cases}$

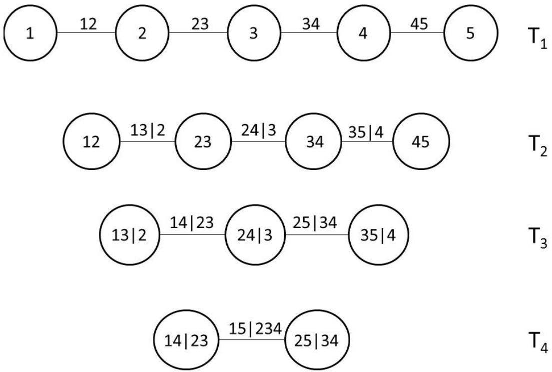
In higher dimensions, even more decompositions are possible: Regular vines

And the indices can be permuted too…

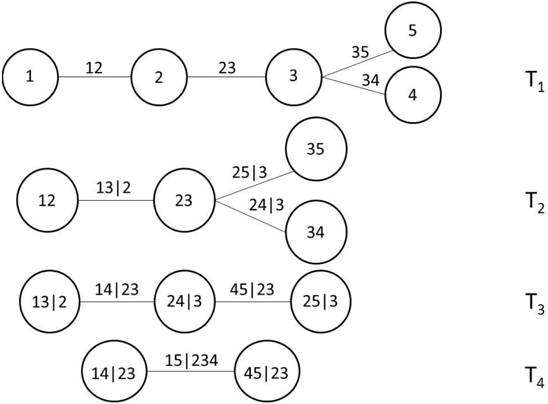
A C-vine in dimension five:
At each level, condition on the same variable



A D-vine in dimension five: Chaining the variables



A non-classified regular vine in dimension five



Vine copulas: Strengths

- Densities are explicit
- Conditioning mechanism also yields simulation algorithms
- Models are easily constructed: any pair copula works
- Highly flexible
  - asymmetries
  - positive/negative dependence
  - tail dependence

Vine copulas: Weaknesses

- Cdf's not explicitly available
- Taking margins destroys the model
- Meaning of chain of simplifying assumptions is not transparent
- Interpretation becomes difficult

Vine copulas: Some literature

Active and fast-moving field. Check out

`http://www-m4.ma.tum.de/forschung/vine-copula-models/`

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics & Economics 44*(2), 182–198.

Bedford, T. and R. M. Cooke (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics 30*(4), 1031–1068.

Brechmann, E. and U. Schepsmeier (2013). Modeling dependence with C- and D-vine copulas: the R package CDVine. *Journal of Statistical Software 52*(3), 1–27.

Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli 19*(2), 462–491.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with fixed marginals and related topics (Seattle, WA, 1993)*, Volume 28 of *IMS Lecture Notes Monogr. Ser.*, pp. 120–141. Hayward, CA: Inst. Math. Statist.