A Primer of Archimedean Copulas in High Dimensions

Johanna Nešlehová
(joint work with Alexander J. McNeil)

Department of Mathematics
ETH Zurich
Switzerland
www.math.ethz.ch/~johanna

Cherry Bud Workshop, March 2007

Why copulas?

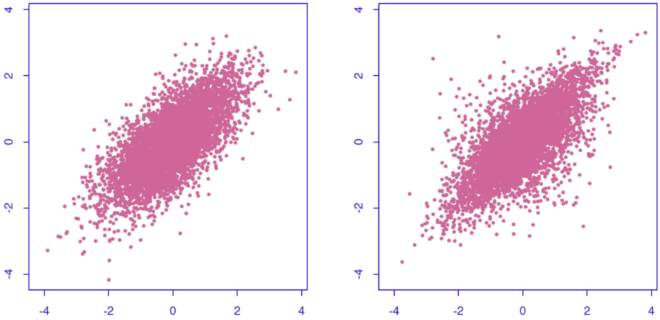
Many useful multivariate distributions exist (normal, t, GH, EVD).

However:

- Danish fire insurance losses over 1 mi. DK from 1980 to 2002. A three dimensional data set consisting of losses for a building and its contents as well as losses of business earnings. Each loss is well modeled by a heavy tailed Generalized Pareto distribution with a different tail index, [McNeil et al].
- Maximum annual flow (in m^3/s) of the Harricana river in Canada and the corresponding volume (in hm^3) for 85 consecutive years, starting 1915 and ending in 1999. For the annual flow, the Gumbel distribution seems appropriate, whereas the annual volume is best modeled by the Gamma distribution, [Genest-Favre].

What to do when marginal distributions differ?

Marginal distributions and a one-number summary of dependence, e.g. Pearson's correlation coefficient do not identify the joint distribution uniquely.



The notion of a copula

Univariate case

Recall from classical statistics:

$$U \sim \mathcal{R}(0,1) \qquad \Rightarrow \qquad F^{-1}(U) \sim F$$

Furthermore, if F is a continuous univariate distribution function,

$$X \sim F \qquad \Rightarrow \qquad F(X) \sim \mathcal{R}(0,1)$$

Higher dimensions

Define a copula C as a joint distribution function (restricted to $[0,1]^d$) whose univariate margins are $\mathcal{R}(0,1)$.

Sklar's theorem for distribution functions

Let H be a d-dimensional joint distribution function with marginals $F_1, \ldots, F_d$. Then there always exists a copula C so that, for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$,

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

If the marginals are continuous then C is unique.

And conversely, if C is a copula and $F_1, \ldots, F_d$ are (arbitrary) univariate marginal distribution functions, then

$$C(F_1(x_1), \ldots, F_d(x_d)) \equiv H(x_1, \ldots, x_d)$$

defines a d-dimensional distribution function with marginals $F_1, \ldots, F_d$.

Sklar's theorem for survival functions

Let $\bar{H}$ be a d-dimensional joint survival function with marginals $\bar{F}_1, \ldots, \bar{F}_d$. Then there always exists a survival copula $\bar{C}$ so that, for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\bar{H}(x_1, \ldots, x_d) = \bar{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)).$$

If the marginals are continuous then $\bar{C}$ is unique.

And conversely, if $\bar{C}$ is a copula and $\bar{F}_1, \ldots, \bar{F}_d$ are (arbitrary) univariate marginal survival functions, then

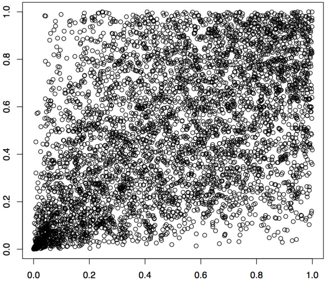
$$\bar{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)) \equiv \bar{H}(x_1, \ldots, x_d)$$

defines a d-dimensional survival function with marginals $\bar{F}_1, \ldots, \bar{F}_d$.

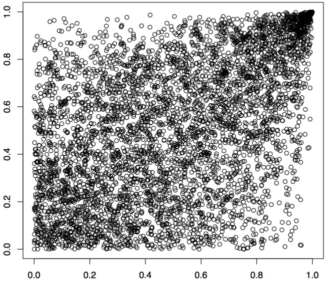
copula

survival copula

$$(U_1, \dots, U_d) \sim C \quad \Leftrightarrow \quad (1 - U_1, \dots, 1 - U_d) \sim \bar{C}$$



copula



survival copula

$$(U_1, \dots, U_d) \sim C \quad \Leftrightarrow \quad (1 - U_1, \dots, 1 - U_d) \sim \bar{C}$$

Copula modeling

- Start with given marginal information in terms of

$$F_1, \ldots, F_d$$

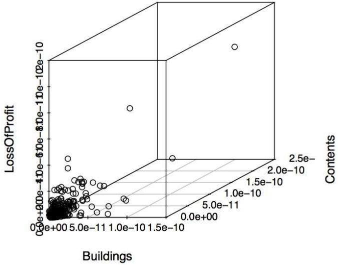
- Any copula C yields a joint model consistent with that information via

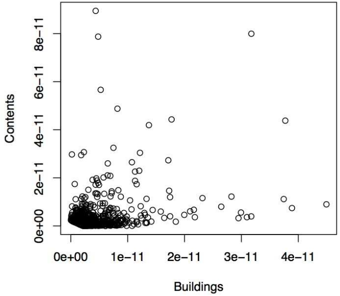
$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

- Data from such a model can be simulated using

$$(U_1, \ldots, U_d) \sim C \quad \Rightarrow \quad (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)) \sim H$$

Danish fire insurance data





Are copulas visible from data?

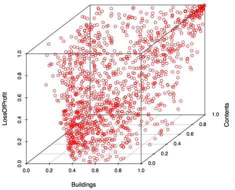
If $F_1, \ldots, F_d$ are continuous then

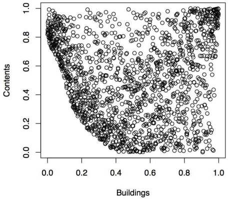
- $(X_1, \ldots, X_d) \sim H \quad \Rightarrow \quad (F_1(X_1), \ldots, F_d(X_d)) \sim C$
- $(X_1, \ldots, X_d) \sim H \quad \Rightarrow \quad (\bar{F}_1(X_1), \ldots, \bar{F}_d(X_d)) \sim \bar{C}$
- $(U_1, \ldots, U_d) \sim C \quad \Leftrightarrow \quad (1 - U_1, \ldots, 1 - U_d) \sim \bar{C}$

For a given data set $\mathbf{X}_1, \ldots, \mathbf{X}_N$

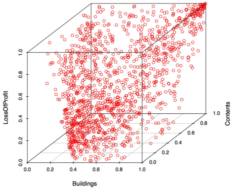
- Construct the component-wise ranks $\mathbf{R}_1, \ldots, \mathbf{R}_N$
- For N large, $\frac{\mathbf{R}_1}{N+1}, \ldots, \frac{\mathbf{R}_N}{N+1}$ yields approximatively a sample from C
- For N large, $1 - \frac{\mathbf{R}_1}{N+1}, \ldots, 1 - \frac{\mathbf{R}_N}{N+1}$ is approximatively a sample from $\bar{C}$

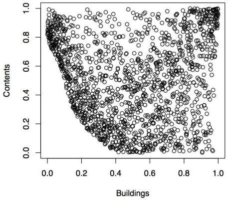
Danish fire insurance data

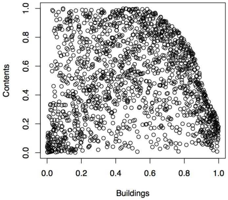




Danish fire insurance data



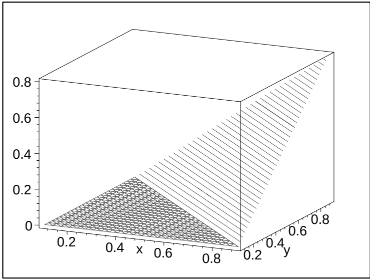


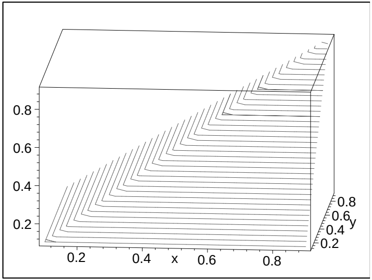


Frechét-Hoeffding bounds

"A copula captures precisely those properties of the joint distribution which are invariant under increasing transformations."

B. Schweizer





$$\max(u_1 + \cdots + u_d - d + 1, 0) \leq C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d)$$

Model selection

Copula modeling does not take away the problem of model choice

Useful copula families

- Exteme value copulas (Gumbel, Galambos, Hüsler-Reiss, ...)
- Elliptical copulas (Gaussian, t, ...)
- Archimedean copulas (Clayton, Gumbel, Frank, ...)
- Other (Plackett, ...)

Helpful techniques for model selection

- Graphical tools
- Goodness of fit tests
- Context

Archimedean copulas

A copula is called Archimedean if it can be written in the form

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

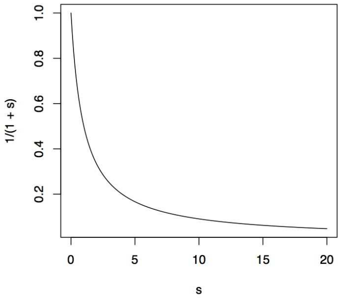
for some generator function ψ and its generalized inverse ψ^{-1}.

The generator ψ satisfies

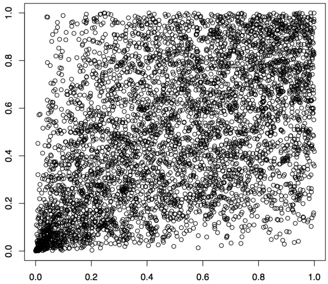
- $\psi : [0, \infty) \to [0, 1]$ with $\psi(0) = 1$ and $\lim_{x\to\infty} \psi(x) = 0$
- ψ is continuous
- ψ is strictly decreasing on $[0, \psi^{-1}(0)]$
- ψ^{-1} is given by $\psi^{-1}(x) = \inf\{u : \psi(u) \leq x\}$

Clayton copula

Take $\psi_\theta(x) = \max\left((1+\theta x)^{-\frac{1}{\theta}}, 0\right)$ for $\theta \geq -\frac{1}{d-1}$.



generator ψ_θ for $\theta = 1$

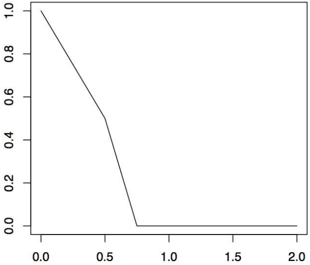


Sample from a Clayton copula for $\theta = 1$

Basic questions

- Is $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$ always well defined?
- What is the interpretation of $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$?
- What are the dependence properties of $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$?
- How can we sample from $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$?
- How can we obtain interesting parametric classes, in particular when $d \geq 3$?

What generators ψ are permissible?



Take

$$\psi(x) = \begin{cases} 1 - x & \text{if } x \in [0, \frac{1}{2}], \\ \frac{3}{2} - 2x & \text{if } x \in [\frac{1}{2}, \frac{3}{4}], \\ 0 & \text{if } x \in [\frac{3}{4}, \infty]. \end{cases}$$

ψ is a generator but $\psi(\psi^{-1}(u_1) + \psi^{-1}(u_2))$ is not a copula.

Necessary and sufficient conditions on ψ for $d = 2$

Ling (1965)

A generator ψ induces a bivariate copula if and only if ψ is convex.

Counterexample for $d \geq 3$

Take $\psi(x) = \max(1 - x, 0)$. Then

$$\psi(\psi^{-1}(u_1) + \psi^{-1}(u_2)) = \max(u_1 + u_2 - 1, 0)$$

which is the Frechét-Hoeffding lower bound. However,

$$\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)) = \max(u_1 + \cdots + u_d - d + 1, 0).$$

Right-hand side is not a copula for $d \geq 3$.

Necessary and sufficient conditions on ψ for $d \geq 2$

Kimberling (1974)

A generator ψ induces an Archimedean copula in any dimension if and only if ψ is completely monotone, i.e. $\psi \in C^\infty(0,\infty)$ and $(-1)^k \psi^{(k)}(x) \geq 0$ for $k = 1, \dots$.

Nelsen ...

A generator ψ induces an Archimedean copula in dimension d if $\psi \in C^d(0,\infty)$ and $(-1)^k \psi^{(k)}(x) \geq 0$ for any $k = 1, \dots, d$.

McNeil & Neslehova (2007)

A generator ψ induces an Archimedean copula in dimension d if and only if ψ is d-monotone, i.e. $\psi \in C^{d-2}(0,\infty)$ and $(-1)^k \psi^{(k)}(x) \geq 0$ for any $k = 1, \dots, d-2$ and $(-1)^{d-2}\psi^{(d-2)}$ is non-negative, non-increasing and convex on $(0,\infty)$.

Example 1

Consider the generator

$$\psi_d^{\mathbf{L}}(x) = \max\left((1-x)^{d-1}, 0\right)$$

- $\psi_2^{\mathbf{L}}$ generates the Frechét-Hoeffding lower bound
- $\psi_d^{\mathbf{L}}$ is k-monotone for $k = 2, \ldots, d$
- $\psi_d^{\mathbf{L}}$ is not k-monotone for $k = d+1, \ldots$
- $\psi_d^{\mathbf{L}}$ can generate an Archimedean copula in dimension up to d but no higher.

Example 2

Consider the Clayton generator

$$\psi_\theta(x) = \max\left((1+\theta x)^{-\frac{1}{\theta}}, 0\right)$$

- ψ_θ is completely monotone for $\theta > 0$
- $\psi_\theta = \exp(-x)$ for $\theta = 0$ which is again completely monotone
- ψ_θ is d-monotone for $\theta \geq -\frac{1}{d-1}$
- ψ_θ is not d-monotone for $\theta < -\frac{1}{d-1}$
- ψ_θ can generate an Archimedean copula in dimension d if and only if $\theta \geq -\frac{1}{d-1}$

Archimedean copulas with completely monotone generators

If ψ is a completely monotonic generator then ψ is the Laplace transform of some non-negative random variable W.

1. Consider an Archimedean copula with completely monotone generator (or LT generator) ψ,

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

2. find W so that ψ is the Laplace transform of W
3. Then C is a survival copula of

$$(X_1, \ldots, X_d) \stackrel{\mathrm{d}}{=} \frac{1}{W}(Y_1, \ldots, Y_d)$$

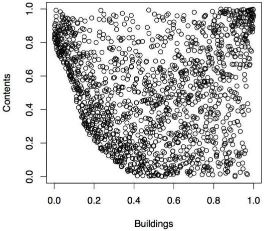
for W and $\mathbf{Y}$ independent and $\mathbf{Y} = (Y_1, \ldots, Y_d)$ a vector of iid standard exponential variables

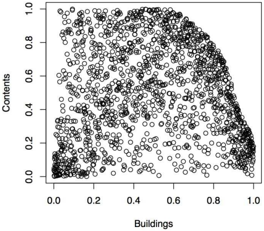
Archimedean copulas with completely monotone generators cont'd

- Archimedean copulas with LT generators have restricted dependence charactreristics

$$\psi \text{ is a LT generator} \Rightarrow \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)) \geq u_1 \cdot \cdots \cdot u_d$$

-





Archimedean copulas and simplex distributions

simplex distributions $\rightsquigarrow$ Archimedean copulas

Consider a non-negative random variable R with $\mathrm{P}(R = 0) = 0$ and a random vector $\mathbf{S}_d$ independent of R and uniformly distributed on

$$\mathcal{S}_d = \left\{ \mathbf{x} \in \mathbb{R}_+^d : |x_1| + \cdots + |x_d| = 1 \right\}$$

Then the survival copula of $\mathbf{X} \stackrel{\mathrm{d}}{=} R\mathbf{S}_d$ is Archimedean.

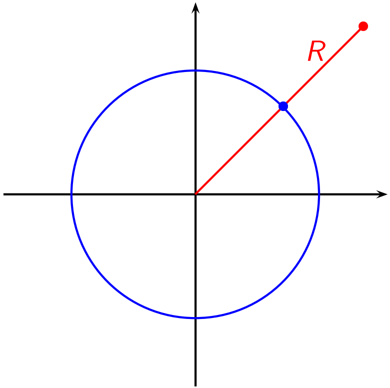
Archimedean copulas $\rightsquigarrow$ simplex distributions

If $C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$ and $\mathbf{U} \sim C$, then

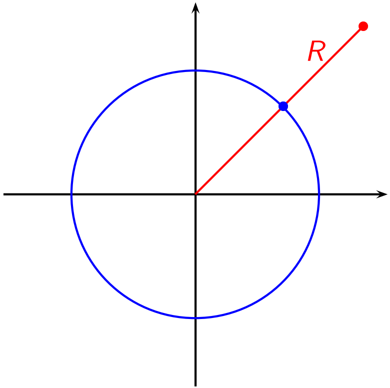
$$\mathbf{X} \stackrel{\mathrm{d}}{=} (\psi^{-1}(U_1), \ldots, \psi^{-1}(U_d))$$

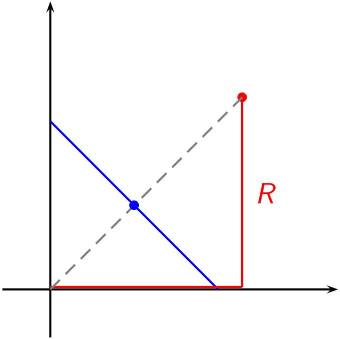
follows a simplex distribution with no atom at zero.

Spherical vs. simplex distributions



Spherical vs. simplex distributions





Reasons from real analysis

- If C is a d-dimensional Archimedean copula, then $\psi(x)$ defines a univariate survival function and

$$\bar{H}(x_1, \ldots, x_d) = \psi(x_1 + \cdots + x_d), \quad (x_1 + \cdots + x_d) \in [0, \infty)^d$$

defines a multivariate survival function with marginals ψ

- If ψ is a d-monotonic generator then ψ is the Williamson d-transform of some non-negative random variable R with $\mathrm{P}(R = 0) = 0$, i.e.

$$\psi(x) = \mathfrak{W}_d F_R(x) = \int_{(x,\infty)} \left(1 - \frac{x}{t}\right)^{d-1} dF_R(t)$$

- Conversely, if R is a non-negative random variable so that $\mathrm{P}(R = 0) = 0$ then its Williamson d-transform generates an Archimedean copula in d dimensions

Archimedean generator vs. radial part

For an Archimedean copula $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$, the radial part of the corresponding simplex distribution has df

$$F_R(x) = 1 - \sum_{k=0}^{d-2} \frac{(-1)^k x^k \psi^{(k)}(x)}{k!} - \frac{(-1)^{d-1} x^{d-1} \psi_+^{(d-1)}(x)}{(d-1)!}$$

For a simplex distribution with radial part R the corresponding survival copula $\psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$ has generator

$$\psi(x) = \mathfrak{W}_d F_R(x) = \int_{(x,\infty)} \left(1 - \frac{x}{t}\right)^{d-1} dF_R(t)$$

Example 1 revisited

For

$$\psi_d^{\mathbf{L}}(x) = \max\left((1-x)^{d-1}, 0\right)$$

the radial part is degenerate, i.e. $R = 1$ a.s. In other words, $\psi_d^{\mathbf{L}}$ generates the survival copula of $\mathbf{S}_d$.

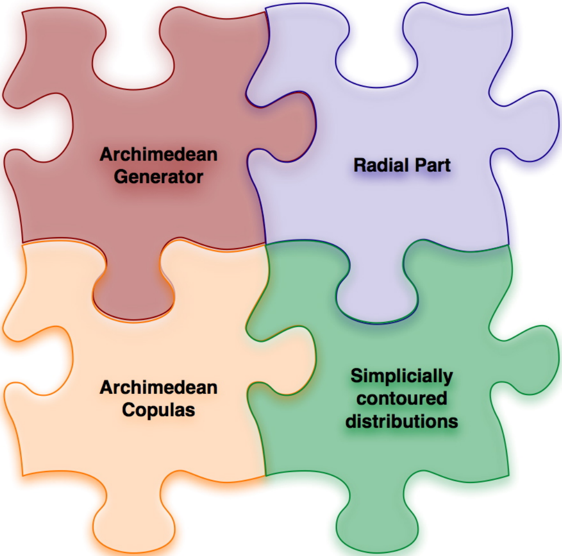
Example 2 revisited

For the Clayton generator

$$\psi_\theta(x) = \max\left((1+\theta x)^{-\frac{1}{\theta}}, 0\right)$$

and $d = 2$,

$$F_R(x) = 1 - (1+\theta x)^{-\frac{1}{\theta}}\left(1 + \frac{x}{1+\theta x}\right)$$


Archimedean Generator
Radial Part
Archimedean Copulas
Simplicially contoured distributions

Sampling from an arbitrary Archimedean copula

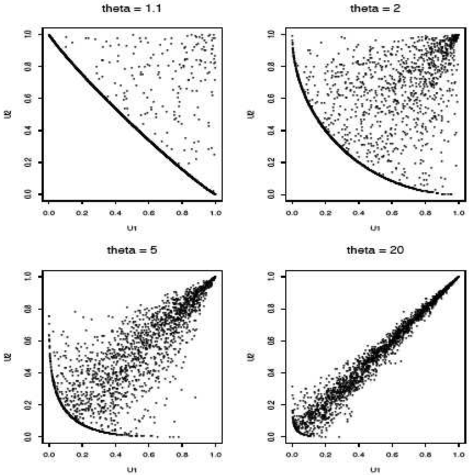
1. Generate R
2. Generate independently $\mathbf{S}_d$ using

$$\mathbf{S}_d \stackrel{\mathrm{d}}{=} \left(\frac{Y_1}{Y_1 + \cdots + Y_d}, \ldots, \frac{Y_d}{Y_1 + \cdots + Y_d} \right)$$

   where $Y_1, \ldots, Y_d$ are iid with $Y_i \sim \mathrm{Exp}(1)$
3. Return

$$\left(\psi \left(R \frac{Y_1}{Y_1 + \cdots + Y_d} \right), \ldots, \psi \left(R \frac{Y_d}{Y_1 + \cdots + Y_d} \right) \right)$$

$$\psi(x) = \max\left((1 - x^{1/\theta}), 0\right), \quad \theta \geq 1$$



A simple goodness-of-fit test

Ingredients

Let C be a d-dimensional Archimedean copula C with generator ψ. Then

$$(U_1, \dots, U_d) \sim C \quad \Rightarrow \quad Y = \psi^{-1}(U_1) + \cdots + \psi^{-1}(U_d) \overset{\mathrm{d}}{=} R$$

$$(U_1, \dots, U_d) \sim C \quad \Rightarrow \quad \mathbf{V} = \left(\frac{\psi^{-1}(U_1)}{Y}, \dots, \frac{\psi^{-1}(U_d)}{Y} \right) \overset{\mathrm{d}}{=} \mathbf{S}_d$$

Numerical tests

- Test whether Y and V_j are independent, $j = 1, \dots, d$
- Test whether $(1 - V_j)^{d-1}$, $j = 1, \dots, d$ are standard uniform

Construction of new families of Archimedean copulas

- Choose a parametric class of non-negative distributions with no atoms at zero

$$\mathcal{R}_\Theta = \{F_\theta : \theta \in \Theta\}$$

- Consider

$$\mathcal{C}_\Theta = \{C_\theta : \theta \in \Theta\}$$

where C_θ, $\theta \in \Theta$ is an Archimedean copula with generator

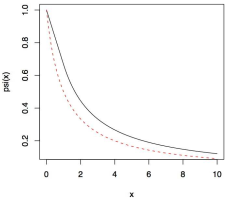
$$\psi_\theta(x) = \mathfrak{W}_d F_\theta(x) = \int_{(x,\infty)} \left(1 - \frac{x}{t}\right)^{d-1} dF_\theta(t)$$

In other words, C_θ, $\theta \in \Theta$ is the survival copula of $\mathbf{X} \stackrel{d}{=} R\mathbf{S}_d$ where $R \sim F_\theta \in \mathcal{R}_\Theta$.

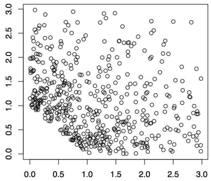
Example 3

Consider $R \sim F_\theta$ corresponding to the Clayton copula and take

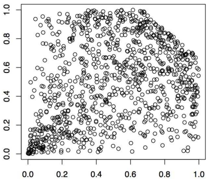
$$R^* \sim F_{\theta,a} \quad \text{where} \quad F_{\theta,a}(x) = 1 - \mathrm{P}(R > x | R > a)$$



$\psi_{\theta,a}$ and ψ_θ



simplex distribution

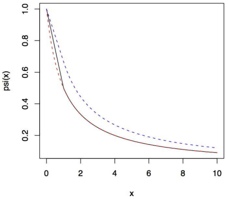


survival copula

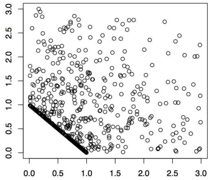
Example 4

Consider $R \sim F_\theta$ corresponding to the Clayton copula and take

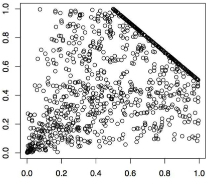
$$\tilde{R} \stackrel{d}{=} \mathbf{1}\{R \leq t\}t + \mathbf{1}\{R > t\}R$$



$\psi_{\theta,t}$, $\psi_{\theta,a}$ and ψ_θ



simplex distribution

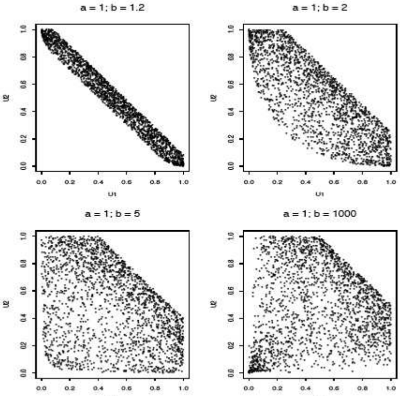


survival copula

Example 5

Consider a radial part R with a density

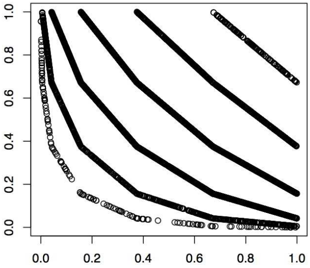
$$f_{a,b}(x) = \frac{ab}{b-a} x^{-2}, \qquad a \leq x \leq b, \quad 0 < a < b$$

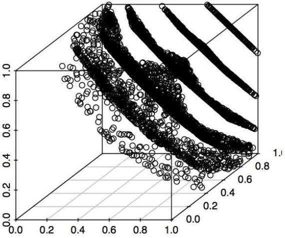


Example 6

Consider a discrete radial part $R \sim F_{n,p}$, $n \in \mathbb{N}$, $p \in [0,1]$:

$$P(R = k) = \binom{n}{(k-1)} p^{k-1}(1-p)^{n-k+1}, \qquad k = 1, \dots, n+1$$





When does an Archimedean copula have a density?

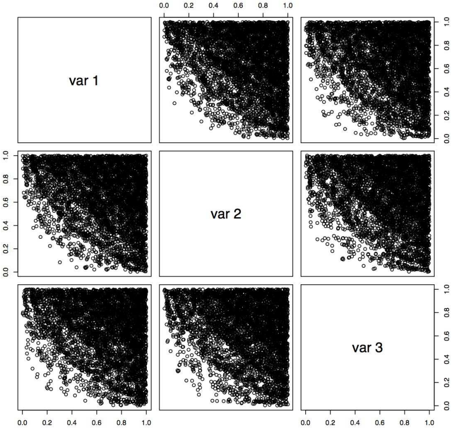
Proposition

Let C be a d-dimensional Archimedean copula with generator ψ and let R denote the radial part of the corresponding simplex distribution. Then

- C has a density if and only if R has a density
- C has a density if and only if $\psi_+^{(d-1)}$ is abs. cont. on $(0, \infty)$
- C has a density whenever ψ generates an Archimedean copula in dimension at least $d+1$.
- If the density exists, then, for almost all $\mathbf{u} \in [0,1]^d$,

$$c(\mathbf{u}) = \frac{\psi^{(d)}(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))}{\psi'(\psi^{-1}(u_1)) \ldots \psi'(\psi^{-1}(u_1))}$$

In particular, all lower dimensional marginals of an Archimedean copula have densities, even if R is purely discrete!



Lower bound

Copulas are not bounded below point-wise

$$\max(u_1 + \cdots + u_d - d + 1, 0) \leq C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d)$$

where the right-hand side is not a copula.

Lower bound

Copulas are not bounded below point-wise

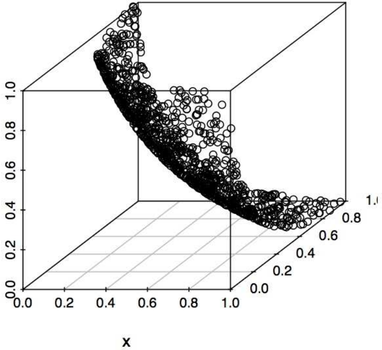
$$\max(u_1 + \cdots + u_d - d + 1, 0) \leq C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d)$$

where the right-hand side is not a copula.

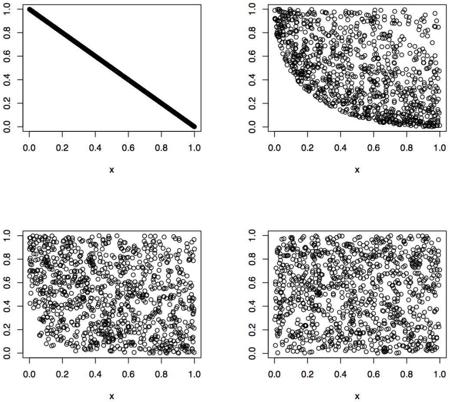
Archimedean copulas are bounded below point-wise

$$\psi_d^{\mathbf{L}}\left((\psi_d^{\mathbf{L}})^{-1}(u_1) + \cdots + (\psi_d^{\mathbf{L}})^{-1}(u_d)\right) \leq \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$$

Lower bound for $d = 3$



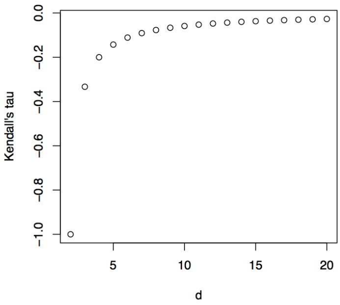
Bivariate marginals of the lower bound for $d = 2, 3, 5, 10$



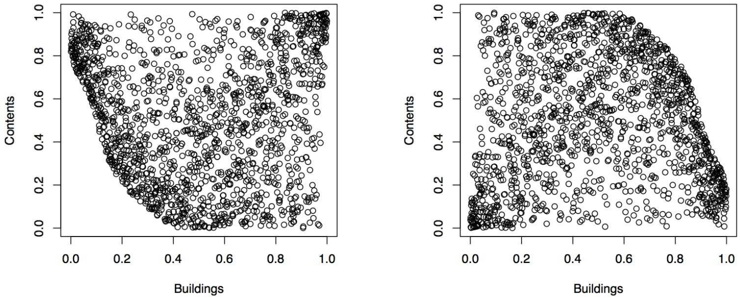
Lower bound on Kendall's tau

For a bivariate margin of a d-dimensional Archimedean copula,

$$\tau = 4\,\mathrm{E}\big(\psi(R)\big) - 1 \quad \text{and} \quad -\frac{1}{2d-3} \leq \tau$$



Conclusions



References I

- Embrechts, P., McNeil, A.J. and Straumann, D. (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls. In: Risk Management: Value at Risk and Beyond, Cambridge University Press.
- Genest, C. and Favre, A.-C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrologic Eng., 12.
- McNeil, A.J., Frey R. and Embrechts, P. (2005) Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.
- McNeil, A.J. and Neslehova, J. (2007) Multivariate Archimedean Copulas, d-monotone Functions and ℓ_1-norm Symmetric Distributions, submitted.

Level sets of Archimedean copulas

Level sets of a copula are

$$L(s) = \left\{\mathbf{u} \in [0,1]^d : C(\mathbf{u}) = s\right\}, \qquad s \in [0,1].$$

For a d-dimensional Archimedean copula:

- $\mathrm{P}^C(L(s)) = \mathrm{P}(R = \psi^{-1}(s))$
- $\mathrm{P}^C(L(0)) = \begin{cases} \frac{(-1)^{d-1}(\psi^{-1}(0))^{d-1}\psi_-^{(d-1)}(\psi^{-1}(0))}{(d-1)!} & \text{if } \psi^{-1}(0) < \infty \\ 0 & \text{otherwise} \end{cases}$
- The level sets are convex