Multivariate Non-Normally Distributed Random Variables
An Introduction to the Copula Approach

Workgroup seminar on climate dynamics
Meteorological Institute at the University of Bonn
18 January 2008, Bonn

Christian Schölzel\textsuperscript{1(2)} and Petra Friederichs\textsuperscript{1}

\textsuperscript{1)} Meteorological Institute at the University of Bonn
Bonn, Germany

\textsuperscript{2)} Laboratoire des Sciences du Climat et l’Environnement (LSCE)
Gif-sur-Yvette, France

\textsuperscript{E2C2: Extreme Events, Causes and Consequences}
**Motivation**
Multivariate non-normally distributed random variables

**Interest**
Estimating multivariate probability density functions

**Problem**
At least one component might be
- precipitation
- wind speed and direction
- relative humidity
- cloud cover
- …
- extremes

**Question**
How to describe random vectors with non-normal marginals?
Background
General problem with multivariate random variables

Standard textbooks


... multivariate normal distribution

Marginal distributions

- numerous parametric descriptions

Dependence

- allows for more complexity
- degrees-of-freedom

... no canonical extension
Related Approaches
How has this problem been addressed so far?

**Multivariate normal distribution**
*) Not an option

**True/native multivariate distribution**
Best option, but often restricted in modeling marginals and dependence

**Mixture models**
Flexible w.r.t. covariance structure, complex in higher dimensions, no account for the original marginal distribution

**Marginal and conditional pdf**
Complexity of the estimation process grows exponentially

**Multivariate KDE**
*) Not a parametric description

**Johnson distribution**
Simple concept, limited dependence structure (precursor & special case)
The copula approach

A brief (and incomplete) historical overview . . .

1949: johnson distribution and other precursors
1959: the term copula has been introdced (Sklar)
2000: applications in econometry
2007: applications in hydrology
Assume a \( m \)-dimensional random vector \( \tilde{X} \) with marginal cumulative distribution functions \( F_{X_1}, \ldots, F_{X_m} \)

**Definition copula**
A copula \( C_{\tilde{X}} \) is a multivariate distribution with standard uniform marginal distributions

**Sklar’s theorem (1959)**
Every joint distribution \( F_{\tilde{X}} \) can be written as a function of its marginal distributions

\[
F_{\tilde{X}}(\tilde{x}) = C_{\tilde{X}}(F_{X_1}(x_1), \ldots, F_{X_m}(x_m))
\]

**Note:** \( C_{\tilde{X}} \) is a copula, since

\[
U_j = F_{X_j}(x_j) \implies U_j \sim \mathcal{U}(0, 1)
\]
**Copula density**
Furthermore, $C_X$ is unique and can be expressed by

$$C_X(u_1, \ldots, u_m) = \int_0^{u_1} \cdots \int_0^{u_m} c_X(u_1', \ldots, u_m') \, du'_1 \cdots du'_m$$

with $u_j = F_{X_j}(x_j)$

**Consequence of Sklar’s theorem**
Every joint probability density can be written as

$$f_X(x) = f_{X_1}(x_1) \cdots f_{X_m}(x_m) \cdot c_X(u_1, \ldots, u_m)$$
Bivariate Examples
Normal-gamma and beta-beta marginals

Christian Schoelzel
Copula Families

Two major classes are **Elliptical copulas** and **Archimedian copulas**

**Elliptical copulas**
Derived from elliptical distributions

- Normal (Gaussian) copula *
- Student’s t-copula

**Archimedian copulas**
Base on so called *generator functions*

\[
C_{\bar{X}}(u_1, \ldots, u_m) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_m))
\]

- Clayton copula: \( \phi_C(u) = u^{-\theta_C} - 1 \)
- Frank copula: \( \phi_F(u) = \log\left(\frac{e^{\theta_F u - 1}}{e^{\theta_F - 1}}\right) \)
- Gumbel copula: \( \phi_G(u) = (-\log u)^{\theta_G} \)
Normal Copulas
Derivation based on Sklar’s theorem

Additional transformation

\[
U_j = F_{\bar{X}}(X_j) \sim \mathcal{U}(0, 1)
\]
\[
Z_j = F_{\mathcal{N}(0, 1)}^{-1}(U_j) \sim \mathcal{N}(0, 1)
\]
\[
\tilde{Z} = (Z_1, \ldots, Z_m)^T \sim \mathcal{N}(0, \Sigma)
\]

Normal copula
Sklar’s theorem leads to the copula

\[
C_{\bar{X}}(u_1, \ldots, u_m) = F_{\mathcal{N}(0, \Sigma)}(F_{\mathcal{N}(0, 1)}^{-1}(u_1), \ldots, F_{\mathcal{N}(0, 1)}^{-1}(u_m))
\]
and the copula density

\[
c_{\bar{X}}(u_1, \ldots, u_m) = \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_m} \cdot C_{\bar{X}}(u_1, \ldots, u_m)
\]
\[
= \frac{f_{\mathcal{N}(0, \Sigma)}(F_{\mathcal{N}(0, 1)}^{-1}(u_1), \ldots, F_{\mathcal{N}(0, 1)}^{-1}(u_m))}{\prod_{j=1}^{m} \left( f_{\mathcal{N}(0, 1)}(F_{\mathcal{N}(0, 1)}^{-1}(u_m)) \right)}
\]
\[ Z_j = F_{\mathcal{N}(0,1)}^{-1}(U_j) \]

\[ U_j = F_{X_j}(X_j) \]

\[ f_{\mathcal{N}(\Sigma,1)}^{-1}(\vec{X}_f) \]

\[ C_{\vec{X}} \]

\[ f_{\vec{X}} \]
**Student t-Copula**

A more flexible case of elliptical copulas

**Student t-Copula**

Based on an extension of the **multivariate t-distribution**

$$C_{\vec{X}} (u_1, \ldots, u_m) = F_{t(\nu, \Sigma)} (F_{t(\nu)}^{-1}(u_1), \ldots, F_{t(\nu)}^{-1}(u_m)),$$

**Tail dependence**

Definition of **upper tail dependence** (lower analogously)

$$\lambda_{\text{up}} = \lim_{u \nearrow 1} P \left( X_1 > F_{X_1}^{-1}(u) \mid X_2 > F_{X_2}^{-1}(u) \right)$$
Different Copulas
Dependence structure and upper/lower tail dependence

- Gaussian
- t-Copula
- t-Copula
- Elliptical family
- Clayton
- Frank
- Gumbel
- Archimedean family
Estimation
The aim is to fit the whole multivariate distribution

**Full maximum likelihood (ML)**
Maximize the log-likelihood functions for the full pdf $f_X$
- consistent estimates for all parameters
- often only numerically feasible

**Inference functions for margins (IFM)**
Two-stage estimation process (1. margins, 2. copula)
- computationally more efficient than ML
- predefined estimators for margins and copula
- no independent parameter estimates

**Canonical maximum likelihood (CML)**
Like IFM but use the empirical CDF of each margin instead
- consistent estimates of the copula parameters
- practical use questionable (see discussion)
Goodness-of-Fit
The probability integral transform (PIT)

**Dimension reduction**
Transform \( \tilde{\mathbf{X}} = (X_1, \ldots, X_m)^T \) to a set of independent, uniform variables

**The probability integral transform (PIT)**
The PIT of \( \tilde{\mathbf{X}} \) is defined as

\[
T_1(X_1) = F_{X_1}(X_1) \\
T_2(X_2) = F_{X_2|X_1}(X_2|X_1) \\
\vdots \\
T_m(X_m) = F_{X_m|X_1,\ldots,X_{m-1}}(X_m|X_1,\ldots,X_{m-1})
\]

**Hypothesis \( \mathcal{H}_0 \)**
\( \tilde{\mathbf{X}} \) comes from the specified multivariate model \( (F_{\tilde{\mathbf{X}}}) \)

\[
Z_j^* = T_j(X_j) \overset{\text{iid}}{\sim} \mathcal{U}(0, 1)
\]
Goodness-of-Fit
The probability integral transform (PIT)

**Dimension reduction**
With the multivariate random variable (above)

\[ Z_j^* = T_j(X_j) \overset{iid}{\sim} \mathcal{U}(0, 1) \]

it follows that

\[ Y^* = \sum_{j=1}^{m} F_{\mathcal{N}(0, 1)}^{-1}(Z_j^*) \]

has a \( \chi^2 \)-distribution with \( m \) degrees of freedom, so that

\[ W^* = F_{\chi^2_m}(Y^*) \]

should be a univariate random variable from \( \mathcal{U}(0, 1) \)

\[ \cdots \text{ univariate GoF tests} \]
Applications
Precipitation and minimum temperature at stations Berlin and Potsdam

Example A: different random variables

Example B: different locations
Multivariate Extremes
How are copulas related to multivariate extremes?

Copulas are related to multivariate extremes

(later)
**Capabilities**

**What are copulas good for?**

**The copula approach**

- Simple and straightforward method to find parametric descriptions of multivariate non-normally distributed random variables

**Characteristics**

- Catch different dependence structures
- Keep proper parametric descriptions of the margins
- Low number of parameters

**Mixture models (alternative and enhancement)**

- More degrees-of-freedom
- Allow only approximative descriptions of the marginal distributions
- Limited in modelling tail dependence
- Based on the assumption of different populations

... mixture models and copulas can be combined
Limitations
What are typical pitfalls?


Limitations

- Limited to the existing copula families
- No general procedure for selecting the copula class

Pitfalls

Questionable practices in finance, risk management, or insurance

- Theoretical value of the copula must not be exaggerated
- There is no dependence separately from the marginal distributions
- Copulas do not solve the problem of dimensionality
Outlook
Where to go?

Outlook
After numerous successful applications in risk management, financial research and more recently in hydrology, it is very likely that copulas will have growing impact in meteorology and climate research
1. **Introduction**
   - Motivation
   - Background and Related Work

2. **The Copula Approach**
   - Basics
   - Copula Families
   - Estimation
   - Goodness-of-Fit

3. **Applications**

4. **Multivariate Extremes**

5. **Discussion**
   - Capabilities
   - Limitations
   - Outlook

►► **Appendix**

Christian Schoelzel