Analyzing dependent data with vine copulas (Lecture 3)

Claudia Czado <cczado@ma.tum.de>, Thomas Nagler and Daniel Kraus
TU München
1. Vine copulas with non parametric pair copulas
   - Motivation
   - Kernel density estimation
   - Vine copulas with non parametric pair copulas

2. Quantile regression
   - Classic linear quantile regression
   - Copula quantile regression
   - D-vine quantile regression

3. Summary
Outline

1 Vine copulas with non parametric pair copulas
   Motivation
   Kernel density estimation
   Vine copulas with non parametric pair copulas

2 Quantile regression
   Classic linear quantile regression
   Copula quantile regression
   D-vine quantile regression

3 Summary
Motivation for non parametric copulas (i)

- We usually assume that a pair-copula $c$ belongs to a parametric family: $c_\theta$, $\theta \in \Theta \subset \mathbb{R}^p$.

- Many families with different characteristics exist:

  Independence
  Gaussian
  Clayton
  Rotated Gumbel

- The parameter $\theta$ can be estimated by series maximum likelihood.

- If the model is correctly specified, the MLE converges to the true parameter and $\hat{c}_\theta \xrightarrow{p} c$. 
Motivation for non parametric pair copulas

What if the parametric assumption is too strong?
MAGIC Gamma Telescope Data Set

- **Source:** archive.ics.uci.edu/ml/datasets/magic+gamma+telescope
- **Variables:**
  - Length: major axis of ellipse
  - Width: minor axis of ellipse
  - Size: 10-log of sum of content of all pixels
  - Conc: ratio of sum of two highest pixels over Size
  - Conc1: ratio of highest pixel over Size
  - Asym: distance from highest pixel to center projected onto major axis
  - M3Long: 3rd root of third moment along major axis
  - M3Trans: 3rd root of third moment along minor axis
  - Alpha: angle of major axis with vector to origin
  - Dist: distance from origin to center of ellipse
- **Classification:**
  - g = gamma (signal): 12332
  - h = hadron (background): 6688
- Data are **MC generated** to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope
- We will consider the dependence among **Length, Width, Size, Asym, Dist.**
Let

- $X_i, i = 1, \ldots, n$ be iid with observed values $x_i$ from density $f$.
- $K$ be a symmetric probability density function on $\mathbb{R}$.

Then, for $b \downarrow 0$, 

$$
\mathbb{E}
\left[
\frac{1}{b} K \left( \frac{X_i - x}{b} \right)
\right] \rightarrow f(x), \quad \text{for all } x \in \mathbb{R} \text{ and } X_i.
$$

Kernel density estimator: (Rosenblatt, 1956; Parzen, 1962)

$$
\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K \left( \frac{x_i - x}{b} \right) \xrightarrow{p} f(x), \quad \text{for all } x \in \mathbb{R}.
$$

$K$ is called the kernel, $b$ is called the bandwidth.
Univariate kernel density estimation

How to select the bandwidth?

- **Idea:** Minimize the mean integrated squared error

\[ \text{MISE} = \int [\hat{f}(x) - f(x)]^2 dx \]

- **Rule of thumb:** \( b = 1.06 \sigma n^{-1/5} \)

  \( \Rightarrow \) asymptotically optimal for Gaussian \( f \) and \( K \).

- There are more sophisticated techniques for non-Gaussian data.
Multivariate kernel density estimation

Let

- $X_i, i = 1, \ldots, n$ be iid random vectors in $\mathbb{R}^d$ with density $f$ and observed values $x_i$.
- $K$ be a symmetric probability density function on $\mathbb{R}^d$.
- $B \in \mathbb{R}^{d \times d}$ invertible bandwidth matrix.

**Multivariate kernel density estimator:**

$$
\hat{f}(x) = \frac{1}{n \det(B)} \sum_{i=1}^{n} K(B^{-1}(x_i - x)) \xrightarrow{p} f(x), \quad \text{for all } x \in \mathbb{R}^d.
$$

**Rule of thumb:** $0.75^{-1/5} \sum_{i=1/2}^{n-1/(4+d)}$. 

Kernel copula density estimation (I)

true density

kernel estimate
Kernel copula density estimation (II)

- Copulas densities have **bounded support**.
- On finite samples, KDE puts probability mass beyond $[0, 1]^2$.

**Transformation trick:**

1. Transform data to normal margins: $Z_1 = \Phi^{-1}(U_1)$, $Z_2 = \Phi^{-1}(U_2)$.

2. Estimate non parametrically the density of $(Z_1, Z_2)$:

   $$f(z_1, z_2) = c(\Phi(z_1), \Phi(z_2))f(z_1)f(z_2).$$

3. **Non parametric copula density estimate**

   $$\hat{c}(u_1, u_2) = \frac{\hat{f}(z_1, z_2)}{\phi(z_1)\phi(z_2)}, \quad z_1 = \Phi^{-1}(u_1), z_2 = \Phi^{-1}(u_2)$$
Kernel copula density estimation (III)

true density

kernel estimate
The curse of dimensionality

Theorem (Stone 1980)

- Let \( \hat{f} \) be a non-parametric kernel density estimator, such that for all densities \( f \) that are 2 times continuously differentiable at \( x \in \mathbb{R}^d \), and some \( r > 0 \),

\[
\hat{f}(x) = f(x) + O_P(n^{-r}).
\]

- Then,

\[
r \leq \frac{2}{4 + d}.
\]

Problem: Convergence slows down as dimension increases.
Theorem (Nagler and Czado 2016)

Let \( f \) is a simplified vine density and \( \hat{f}_{\text{vine}} \) is a density estimator, where the pair copulas and conditional distribution functions are non parametrically estimated. Then under regularity conditions,

\[
\hat{f}_{\text{vine}}(x) = f(x) + O_p(n^{-1/3}),
\]

where \( r \) does not depend on \( d \).

\[\Rightarrow\] There is no curse of dimensionality!
Nagler et al. (2017) showed that the transformation local likelihood (TTL) kernel estimator for pair copulas performs better than multivariate (penalized) Bernstein and B-spline copula estimators under stronger dependence.

TTL uses for $f$ the local polynomial likelihood estimator of Loader (2006) and later applied for bivariate copulas by Geenens et al. (2017).

Nagler and Czado (2016) showed no curse of dimensionality in simplified vines.
conditional quantiles of a response $Y$ given the covariate values $x_i, i = 1, \ldots, d$ are needed for prediction.

$$F_{Y|X}^{-1}(\alpha|x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i$$

linearity assumption violated under non Gaussian dependence, therefore quantile crossing possible (Bernard and Czado, 2015)
Classic linear quantile regression

conditional quantiles of a response $Y$ given the covariate values $x_i, i = 1, \ldots, d$ are needed for prediction.

Classic linear quantile regression

$$F_{Y|X}(\alpha|x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i$$

linearity assumption violated under non Gaussian dependence, therefore quantile crossing possible (Bernard and Czado, 2015)
Desired extensions

- Want a method without quantile crossings.
- Want to accommodate different non Gaussian dependence structures using copulas.
- Want copula model where we can easily compute conditional densities and quantiles.
- Want to have an automatic variable selection method.
- Want to be able to handle discrete response or covariates.
## Copula quantile regression

**Original scale**

<table>
<thead>
<tr>
<th>Response variable</th>
<th>( Y \sim F_Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor variables</td>
<td>( X = (X_1, \ldots, X_d) ), ( X_j \sim F_{X_j} )</td>
</tr>
</tbody>
</table>

### Quantile regression

\[
F^{-1}_{\frac{Y}{X}}(\alpha|\mathbf{x}) = \quad .
\]
### Copula quantile regression

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Original scale</th>
<th>Copula scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$ ~ $F_Y$</td>
<td>$V = F_Y(Y)$</td>
<td></td>
</tr>
<tr>
<td>$X = (X_1, \ldots, X_d)$, $X_j ~ F_{X_j}$</td>
<td>$U = (U_1, \ldots, U_d)$, where $U_j = F_{X_j}(X_j)$</td>
<td></td>
</tr>
</tbody>
</table>

**Quantile regression**

$$F_{Y|X}^{-1}(\alpha|x) = \text{.}$$
Copula quantile regression

Response variable

- $Y \sim F_Y$

Predictor variables

- $X = (X_1, \ldots, X_d)$,
- $X_j \sim F_{X_j}$

Original scale

Copula scale

- $V = F_Y(Y)$
- $U = (U_1, \ldots, U_d)$,
  where $U_j = F_{X_j}(X_j)$

Copula quantile regression

$$F_{Y|X}^{-1}(\alpha|x) = F_Y^{-1} \left( C_{V|U}^{-1}(\alpha|u) \right)$$

- Here $C_{V,U}$ denotes the joint copula of $(V, U)$ and $C_{V|U}$ the conditional distribution of $V$ given $U$. 

.  

21 / 32
Copula quantile regression

Original scale

Response variable

\( Y \sim F_Y \)

Predictor variables

\( X = (X_1, \ldots, X_d), \)

\( X_j \sim F_{X_j} \)

Copula scale

\( V = F_Y(Y) \)

\( U = (U_1, \ldots, U_d), \)

where \( U_j = F_{X_j}(X_j) \)

Copula quantile regression

\[
\hat{F}^{-1}_{Y|X}(\alpha|x) = \hat{F}^{-1}_Y\left(\hat{C}^{-1}_{V|U}(\alpha|u)\right)
\]

- Here \( C_{V,U} \) denotes the joint copula of \((V, U)\) and \( C_{V|U} \) the conditional distribution of \( V \) given \( U \).

- For estimation we model marginals \( F_Y \) and \( F_{X_j} \) non parametrically and \( \hat{C}_{V,U} \) is an estimated D-vine copula.
D-vine copulas

- Model dependence **pairwise** (with conditioning)
- Tree representation: each edge corresponds to a copula
- D-vine copula: each tree is a path
D-vine copulas

- Density at \( u = (u_1, \ldots, u_4) \) factorizes:

\[
    c_{1234}(u) = c_{12}(u_1, u_2) \times c_{23}(u_2, u_3) \times c_{34}(u_3, u_4) \quad \text{(tree 1)}
\]
\[
    \times c_{13;2}(u_{1|2}, u_{3|2}) \times c_{24;3}(u_{2|3}, u_{4|3}) \quad \text{(tree 2)}
\]
\[
    \times c_{14;23}(u_{1|23}, u_{4|23}), \quad \text{(tree 3)}
\]

where \( u_{i|D} = C_{i|D}(u_i | u_D) \).

- Arguments computed recursively via h-functions:

\[
    u_{i|D \cup j} = h_{i|j;D}(u_{i|D}, u_{j|D}) = \frac{\partial C_{ij;D}(u_{i|D}, u_{j|D})}{\partial u_{j|D}}
\]

- Special cases: Gaussian copulas, t copula, Clayton copula
A D-vine copula is used since there are closed form expressions for conditional distributions and densities as well as conditional quantiles. For \( d = 4 \) we have

**conditional density:**

\[
c_{1|234} = c_{12} \times c_{1,3;2} \times c_{1,4;23}
\]

**conditional distribution:**

\[
C_{1|234}(u_1|u_2, u_3, u_4) = h_{1|4;23}(u_1|234, u_4|234)
\]

**conditional quantile:**

\[
C^{-1}_{1|234}(\alpha | u_2, u_3, u_4) = h_{1|2}^{-1}\left[h_{1|3;2}^{-1}\left\{ h_{1|4;23}^{-1}(\alpha, u_4|23), u_3|2 \right\}, u_2 \right]
\]
Kraus and Czado (2017) propose a **forward variable selection algorithm** based on the

**conditional log-likelihood**

\[ cll(D) = \sum_{i=1}^{n} \ln c_{1\mid2,\ldots,D}(\hat{u}_{i,1}, \ldots, \hat{u}_{i,D}) \quad (+\text{penalty}), \]

where \( \hat{u}_{i,1} := \hat{F}_{j}(y_{i}) \) and \( \hat{u}_{i,j} := \hat{F}_{j}(x_{ij}) \) for \( j = 2, \ldots, D \).

1. Fix \( Y \) as first variable.
2. Check which covariate increases \( cll \) the most.
   - If no improvement → stop.
   - If improvement → add covariate and continue with 2.
Extensions to D-vine quantile regression

- **Non parametric pair copulas**
  - Different non parametric pair copula estimation methods studied in Nagler (2014) and Nagler et.al (2017).
  - Nagler and Czado (2016) showed no curse of dimensionality in simplified vines.

- **Allowing for discrete variables**
  - Panagiotelis et al. (2012), Stöber (2013) and Stöber et al. (2015) showed how to handle discrete variables in vines.
  - Nagler (2018) developed valid non parametric jittering method.
  - Schallhorn et al. (2017) developed a D-vine quantile regression approach for mixed discrete variables using these approaches.
Some more details

- **Nonparametric pair copula estimation**
  - It replaces the standard kernel estimator by a local polynomial approximation (cf. Loader 1999)

- **Non parametric jittering in copulas (Nagler 2017)**
  - pmf of the discrete variable = pdf jittered continuous variable at the observed values.
  - provides consistent non parametric estimates in contrast to parametric jittering for copulas (Nikoloulopoulos 2013)
Outline

1. Vine copulas with non parametric pair copulas
   - Motivation
   - Kernel density estimation
   - Vine copulas with non parametric pair copulas

2. Quantile regression
   - Classic linear quantile regression
   - Copula quantile regression
   - D-vine quantile regression

3. Summary
Summary

Non parametric estimation:

- Introduced non parametric kernel estimation.
- Considered non parametric pair copula estimation using kernels.
- Showed no curse of dimensionality for $d$ dimensional density estimation for simplified vine densities.

Quantile regression:

- Introduced classic linear quantile regression.
- Considered copula based extensions of quantile regression.
- Introduced D-vine quantile regression.
- Developed automatic forward selection of covariates.
References

Conditional quantiles and tail dependence.  

Probit transformation for nonparametric kernel estimation of the copula density.  
*Bernoulli* 23(3), 1848–1873.

D-vine copula based quantile regression.  
*Computational Statistics & Data Analysis* 110C, 1–18.

*Local regression and likelihood.*  

Kernel methods for vine copula estimation.  
Masterarbeit, Technische Universität München.

A generic approach to nonparametric function estimation with mixed data.  

Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas.  
*Journal of Multivariate Analysis* 151, 69–89.


