Analyzing dependent data with vine copulas (Lecture 3)

Claudia Czado <cczado@ma.tum.de>, Thomas Nagler and Daniel Kraus

TU München

École d'été 2018 Sept 2-5,2018

Overview

Outline

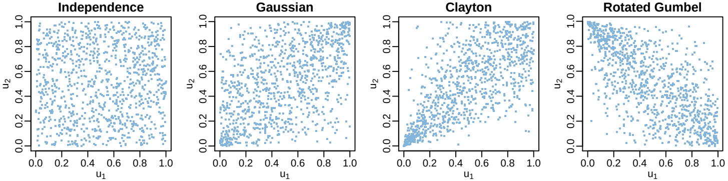
1 Vine copulas with non parametric pair copulas
- Motivation
- Kernel density estimation
- Vine copulas with non parametric pair copulas

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

3 Summary

Motivation for non parametric copulas (i)

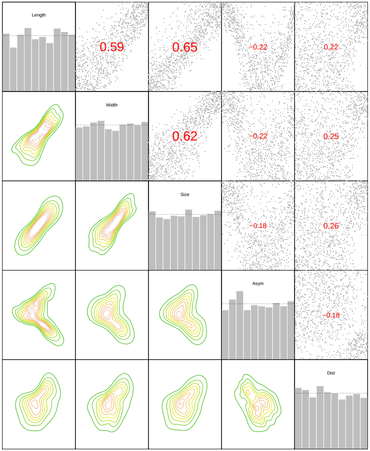
- We usually assume that a pair-copula c belongs to a parametric family: c_θ, $\theta \in \Theta \subset \mathbb{R}^p$.
- Many families with different characteristics exist:



- The parameter θ can be estimated by series maximum likelihood.
- If the model is correctly specified, the MLE converges to the true parameter and $\widehat{c}_\theta \xrightarrow{p} c$.

Motivation for non parametric pair copulas (I)

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.

Univariate kernel density estimation

- 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 $b \searrow 0$,

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

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

:

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

- 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

$$\mathrm{MISE} = \int \left[\widehat{f}(x) - f(x)\right]^2 dx$$

- **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

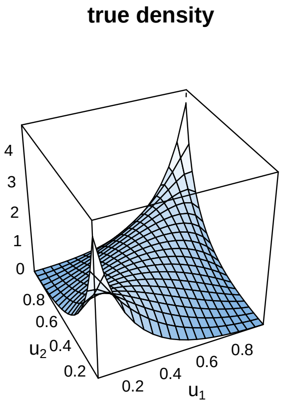
- $\boldsymbol{X}_i$, $i = 1, \ldots, n$ be *iid* random vectors in $\mathbb{R}^d$ with density f and observed values $\boldsymbol{x}_i$.
- K be a symmetric probability density function on $\mathbb{R}^d$.
- $B \in \mathbb{R}^{d \times d}$ invertible bandwidth matrix.

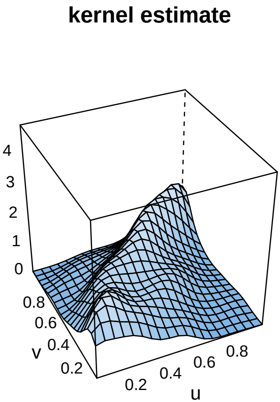
Multivariate kernel density estimator:

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

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

Kernel copula density estimation (I)





Kernel copula density estimation (II)

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

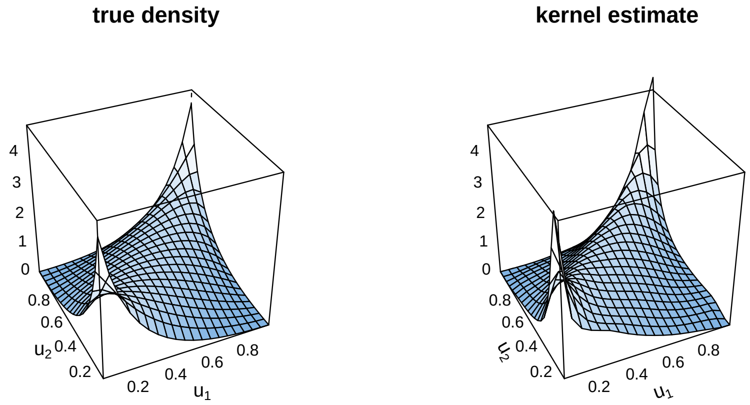
  1. Transform data to normal margins: $Z_1 = \Phi^{-1}(U_1)$, $Z_2 = \Phi^{-1}(U_2)$.
  2. Estimate non parametrically the density of (Z_1, Z_2):

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

3. **Non parametric copula density estimate**

$$\widehat{c}(u_1, u_2) = \frac{\widehat{f}(z_1, z_2)}{\phi(z_1)\phi(z_2)}, \qquad z_1 = \Phi^{-1}(u_1), z_2 = \Phi^{-1}(u_2)$$

Kernel copula density estimation (III)



The curse of dimensionality

Theorem (Stone 1980)

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

$$\widehat{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + O_P(n^{-r}).$$

- *Then,*

$$r \leq 2/(4 + d).$$

Problem: Convergence slows down as dimension increases.

Vines with non parametric pair copulas

Theorem (Nagler and Czado 2016)

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

$$\widehat{f}_{vine}(\boldsymbol{x}) = f(\boldsymbol{x}) + O_p(n^{-1/3}),$$

where r does not depend on d.

$\Rightarrow$ There is no curse of dimensionality!

Non parametric pair copula methods

- 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.

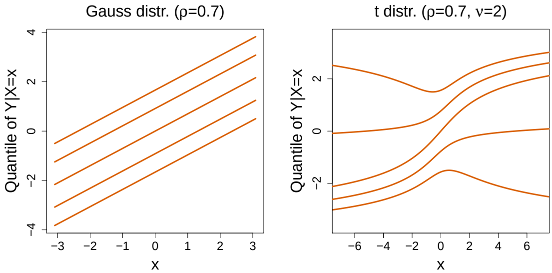
Classic linear quantile regression

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

Classic linear quantile regression

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

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



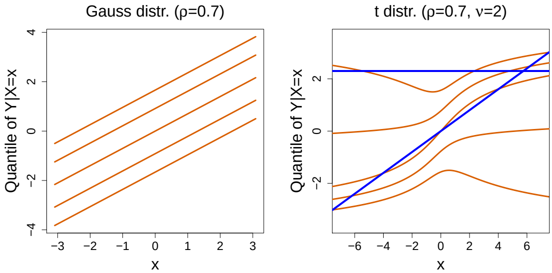
Classic linear quantile regression

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

Classic linear quantile regression

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

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 accomodate 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
Response variable	$Y \sim F_Y$
Predictor variables	$\mathbf{X} = (X_1, \ldots, X_d)$,
	$X_j \sim F_{X_j}$

quantile regression

$$F^{-1}_{Y|\mathbf{X}}(\alpha|\mathbf{x}) =$$

Copula quantile regression

	Original scale	Copula scale
Response variable	$Y \sim F_Y$	$V = F_Y(Y)$
Predictor variables	$\mathbf{X} = (X_1, \ldots, X_d)$,	$\mathbf{U} = (U_1, \ldots, U_d)$,
	$X_j \sim F_{X_j}$	where $U_j = F_{X_j}(X_j)$

quantile regression

$$F^{-1}_{Y|\mathbf{X}}(\alpha|\mathbf{x}) =$$

.

Copula quantile regression

	Original scale	Copula scale
Response variable	$Y \sim F_Y$	$V = F_Y(Y)$
Predictor variables	$\mathbf{X} = (X_1, \dots, X_d)$,	$\mathbf{U} = (U_1, \dots, U_d)$,
	$X_j \sim F_{X_j}$	where $U_j = F_{X_j}(X_j)$

Copula quantile regression

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

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

Copula quantile regression

	Original scale	Copula scale
Response variable	$Y \sim F_Y$	$V = F_Y(Y)$
Predictor variables	$\mathbf{X} = (X_1, \dots, X_d)$,	$\mathbf{U} = (U_1, \dots, U_d)$,
	$X_j \sim F_{X_j}$	where $U_j = F_{X_j}(X_j)$

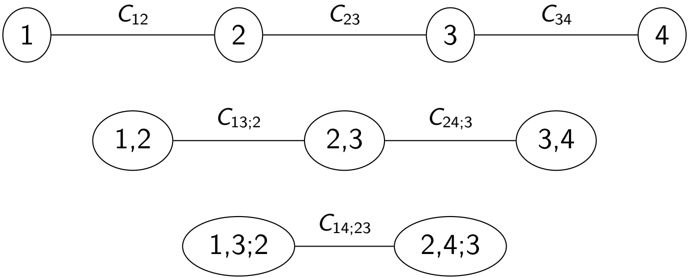
Copula quantile regression

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

- Here $C_{V,\mathbf{U}}$ denotes the joint copula of $(V, \mathbf{U})$ and $C_{V|\mathbf{U}}$ the conditional distribution of V given $\mathbf{U}$.
- For estimation we model marginals F_Y and F_{X_j} non parametrically and $\hat{C}_{V,\mathbf{U}}$ 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 $\mathbf{u} = (u_1, \ldots, u_4)$ factorizes :

$$
\begin{aligned}
c_{1234}(\boldsymbol{u}) = \;& c_{12}(u_1, u_2) \times c_{23}(u_2, u_3) \times c_{34}(u_3, u_4) && \text{(tree 1)}\\
& \times c_{13;2}(u_{1|2}, u_{3|2}) \times c_{24;3}(u_{2|3}, u_{4|3}) && \text{(tree 2)}\\
& \times c_{14;23}(u_{1|23}, u_{4|23}), && \text{(tree 3)}
\end{aligned}
$$

where $u_{i|D} = C_{i|D}(u_i|\boldsymbol{u}_D)$.

- 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

D-vine quantile regression

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:

$$C^{-1}_{1|234}(\alpha \mid u_2, u_3, u_4) = h^{-1}_{1|2}\left[h^{-1}_{1|3;2}\left\{h^{-1}_{1|4;23}(\alpha, u_{4|23}), u_{3|2}\right\}, u_2\right]$$

Forward selection in D-vine regression

Kraus and Czado (2017) propose a forward variable selection algorithm based on the

conditional log-likelihood

$$cll(D) = \sum_{i=1}^{n} \ln c_{1|2,\ldots,D}(\hat{u}_{i,1}|,\ldots,\hat{u}_{i,D}) \qquad (+penalty),$$

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 $\rightarrow$ stop.
   - If improvement $\rightarrow$ 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**
  - Nagler (2014) applies the transformation estimator of Geenens et. al (2014).
  - 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

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

Bernard, C. and C. Czado (2015).
Conditional quantiles and tail dependence.
Journal of Multivariate Analysis 138, 104–126.

Geenens, G., A. Charpentier, D. Paindaveine, et al. (2017).
Probit transformation for nonparametric kernel estimation of the copula density.
Bernoulli 23(3), 1848–1873.

Kraus, D. and C. Czado (2017).
D-vine copula based quantile regression.
Computational Statistics & Data Analysis 110C, 1–18.

Loader, C. (2006).
Local regression and likelihood.
Springer Science & Business Media.

Nagler, T. (2014).
Kernel methods for vine copula estimation.
Masterarbeit, Technische Universität München.

Nagler, T. (2018).
A generic approach to nonparametric function estimation with mixed data.
Statistics & Probability Letters 137, 326–330.

Nagler, T. and C. Czado (2016).
Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas.
Journal of Multivariate Analysis 151, 69–89.

References

Nagler, T., C. Schellhase, and C. Czado (2017).
Nonparametric estimation of simplified vine copula models: comparison of methods.
Dependence Modeling 5, 99–120.

Nikoloulopoulos, A. K. (2013).
On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood.
Journal of Statistical Planning and Inference 143(11), 1923–1937.

Panagiotelis, A., C. Czado, and H. Joe (2012).
Pair copula constructions for multivariate discrete data.
Journal of the American Statistical Association 107(499), 1063–1072.

Schallhorn, N., D. Kraus, T. Nagler, and C. Czado (2017).
D-vine quantile regression with discrete variables.
arXiv preprint, arXiv:1705.08310.

Stöber, J. (2013).
Regular Vine Copulas with the simplifying assumption, time-variation, and mixed discrete and continuous margins.
Dissertation, Technische Universität München, München.

Stöber, J., H. G. Hong, C. Czado, and P. Ghosh (2015).
Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses.
Computational Statistics & Data Analysis 88, 28–39.

Stone, C. J. (1980).
Optimal rates of convergence for nonparametric estimators.
The annals of Statistics, 1348–1360.