

# On the Choice of Parametric Families of Copulas

Radu Craiu

Department of Statistics  
University of Toronto

Collaborators:

Elif Acar, University of Toronto  
Mariana Craiu, University Politehnica, Bucharest  
Fang Yao, University of Toronto

Vienna, July 2008

# Outline

- 1 Brief Review of Copulas
  - What is a Copula and Why should we care?
- 2 Copula misspecification
  - Simulation study of the effects of copula misspecification
- 3 Choice of a Copula Family
  - A nonparametric estimate of distributional distances

# Copulas

- Copulas present one possible approach to model dependence.
- If  $X$ ,  $Y$  are continuous random variables with distribution functions (df)  $F_X$  and, respectively,  $F_Y$  we specify the joint df using the copula  $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$  such that

$$F_{XY}(F_X^{-1}(u), F_Y^{-1}(v)) = \Pr(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(v)) = C(u, v).$$

- The copula  $C$  bridges the marginal distributions of  $X$  and  $Y$ . Interesting: connection between dependence structures and various families of copulas.
- Popular class: *Archimedean copulas*

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)),$$

where  $\phi$  is a continuous, strictly decreasing function  $\phi : [0, 1] \rightarrow [0, \infty]$  and

$$\phi^{[-1]} = \begin{cases} \phi^{-1}(t) & \text{if } 0 \leq t \leq \phi(0) \\ \phi(0) & \text{if } \phi(0) \leq t \leq \infty. \end{cases}$$

# Copulas (cont'd)

- Examples:

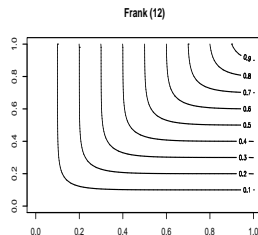
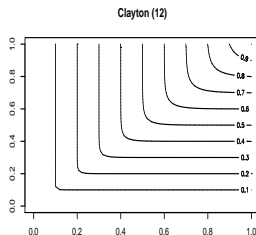
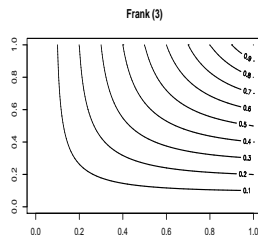
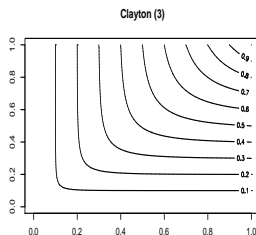
Clayton's copula:  $C(u, v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-1/\theta}$ .

Frank's copula:  $C(u, v) = -\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]$ .

- For the purpose of inference, **given a family of copulas has been selected**, of interest is the estimation of  $\theta$  as well as the marginal distributions' parameters, say  $\lambda_X, \lambda_Y$ .
- The effect of marginal models misspecification has been well documented. Also important is the effect of copula misspecification, especially when of interest are conditional estimates such as  $E[X|Y = y], \text{Var}(X|Y = y)$ .
- Central to the performance of the model is the correct specification of the copula family.

# Copulas (cont'd)

Contour plots of the bivariate cdf:



# Copula Misspecification: A simulation study

- We assume that the marginals are known.
- We generate data following the bivariate Clayton's density.
- We fit a model using **Frank's copula**. We are interested in evaluating the bias for conditional mean and variance estimators.
- Each simulation study has a sample size of  $n = 500$  and we replicate each study  $K = 200$  times.
- The conditional means are computed via Monte Carlo using a sample of size  $M = 5000$ .

## Simulation Results

Clayton's $\theta = 3$ ; $F_X = \text{Exp}(2)$ , $F_Y = \text{Exp}(1)$				
$y_0$	0.5	1.0	1.5	2.5
$B(\mu_{y_0})$	-0.067 (0.009)	-0.072 (0.014)	-0.003 (0.022)	0.140 (0.037)
$B(\sigma_{y_0}^2)$	0.142 (0.026)	0.364 (0.043)	0.646 (0.080)	1.041 (0.147)
Clayton's $\theta = 3$ ; $F_X = F_Y = \text{Weibull}(1, 2)$				
$y_0$	0.5	1.0	1.5	2.5
$B(\mu_{y_0})$	-0.052 (0.042)	-0.285 (0.048)	-0.357 (0.051)	-0.170 (0.071)
$B(\sigma_{y_0}^2)$	-0.061(0.018)	-0.647 (0.209)	-1.036 (0.279)	-1.030 (0.400)
Clayton's $\theta = 12$ ; $F_X = F_Y = \text{Weibull}(1, 2)$				
$y_0$	0.5	1.0	1.5	2.5
$B(\mu_{y_0})$	0.011 (0.012)	-0.008(0.016)	-0.035 (0.023)	-0.118 (0.047)
$B(\sigma_{y_0}^2)$	0.056 (0.006)	0.076 (0.014)	0.050 (0.043)	-0.294 (0.095)

# Outline of the approach proposed

- **Problem:** Given a sample  $\{x_i, y_i\}_{1 \leq i \leq n}$  choose the family of copulas that best approximates the true unknown joint density  $c^*(x, y)$ .
- Assume marginals are known and (without loss of generality)  $\text{Uniform}(0, 1)$ .
- Compute a nonparametric estimate of the two-dimensional density.
- Among a set of possible families find the one who is closest (wrt a certain distributional distance) to the nonparametric estimate.
- Compare two different discrepancies: Kullback-Leibler and Hellinger.

# Nonparametric Estimate

- A sample of size  $n$  from  $c^*$ :  $\{(u_i, v_i) \in [0, 1]^2 : 1 \leq i \leq n\}$ .
- The kernel density is defined by  
$$\hat{c}^*(x; H) = n^{-1} \sum_{i=1}^n K_H(x - X_i),$$
 where  $x = (x_1, x_2)^T$ ,  
 $X_i = (u_i, v_i)$  and  $K_H(x) = |H|^{-1/2} K(H^{-1/2}x)$ .
- $H$  is non-diagonal since there is a large probability mass oriented away from the coordinate directions
- $H$  is data-driven (least squares cross-validation).

# Distributional Distances

- Kullback-Leibler discrepancy is defined as

$$KL(f, g) = \int \log(f(x)/g(x))f(x)dx.$$

- The Hellinger distance is

$$HE^2(f, g) = \int f(x) \left[ 1 - \frac{\sqrt{g(x)}}{\sqrt{f(x)}} \right]^2 dx.$$

# Computing the distance

- Two families of copula densities  $\mathcal{A} = \{c_\alpha : \alpha \in A\}$  and  $\mathcal{B} = \{c_\beta : \beta \in B\}$ , where  $\alpha$  and  $\beta$  are copula parameters.
- Find the MLE's  $\hat{\alpha}$  and  $\hat{\beta}$ .
- Generate a sample  $\{(\tilde{u}_i, \tilde{v}_i) : 1 \leq i \leq m\}$  drawn from  $c_{\hat{\alpha}}$
- Compute

$$\widehat{KL}(c_{\hat{\theta}}, \hat{c}^*) = \frac{1}{m} \sum_{i=1}^m c_{\hat{\theta}}(\tilde{u}_i, \tilde{v}_i) [\log(c_{\hat{\theta}}(\tilde{u}_i, \tilde{v}_i)) - \log(\hat{c}^*(\tilde{u}_i, \tilde{v}_i))],$$

$$\theta = \alpha, \beta.$$

- Similarly for the Hellinger distance:

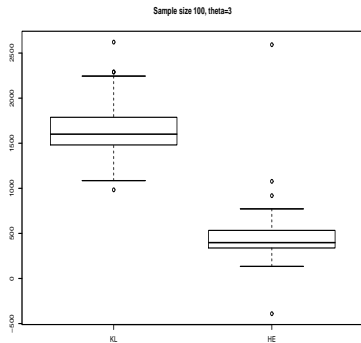
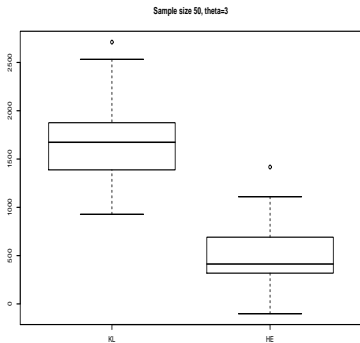
$$\widehat{HE}^2(c_{\hat{\theta}}, \hat{c}^*) = \frac{1}{m} \sum_{i=1}^m \left[ 1 - \frac{\sqrt{\hat{c}^*(\tilde{u}_i, \tilde{v}_i)}}{\sqrt{c_{\hat{\theta}}(\tilde{u}_i, \tilde{v}_i)}} \right]^2, \quad \theta = \alpha, \beta.$$

# Simulation Results

Method \ $n$	50	100	300	500
Clayton's $\theta = 3$				
KL	100	100	100	100
HE	99	99	100	100
Clayton's $\theta = 8$				
KL	100	100	100	100
HE	100	100	100	100
Clayton's $\theta = 12$				
KL	100	100	100	100
HE	100	100	100	100

# Further Comparison

Compare difference in distances measured by KL and HE ( $\theta = 3$ ).



# Further Comparison

Difference in distances measured by KL and HE ( $\theta = 8, 12$ ).

