Multivariate Association Measures: A Review

Vojtěch Kika

KPMS, MFF UK
Department of Mathemathics, KU Leuven

vojtech.kika@kuleuven.be

ROBUST 2018, Rybník, CZ

Thursday 25$^{\text{th}}$ January, 2018

Contents

Motivation

- Having a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$, how would one simply express strength of association between its components?
- We know that for $d = 2$.
  - Kendall's Tau
  - Spearman's Rho
  - Blomqvist's Beta
  - Gini's Gamma
  - ...
- Generalize these!

Setting & Notation

- F distribution function of $\boldsymbol{X}$
- X_i with continuous marginal distribution function F_i, for $i = 1, \dots, d$
- There exists a unique copula function $C : [0,1]^d \to [0,1]$ satisfying

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

- $C(\boldsymbol{u}) = \mathrm{P}(\boldsymbol{U} \leq \boldsymbol{u})$ where $U_i = F_i(X_i)$ for $i = 1, \dots, d$.
- Survival function $\overline{C}(\boldsymbol{u}) = \mathrm{P}(\boldsymbol{U} > \boldsymbol{u})$.
- Independence copula $\Pi(\boldsymbol{u}) = \prod_{i=1}^d u_i$,
- Comonotonicity copula $M(\boldsymbol{u}) = \min\{u_i; 1 \leq i \leq d\}$.

Generalization of a Bivariate Measures

- $\kappa_2(X_i, X_j)$ given bivariate association measure between X_i and X_j
- d-variate pairwise (PW) association measure

$$\kappa_d(X_1, \ldots, X_d) = \frac{1}{\binom{d}{2}} \sum_{\substack{i,j=1 \\ i \neq j}}^{d} \kappa_2(X_i, X_j)$$

- Copula-based approach
  - Express $\kappa_2(X_i, X_j)$ as a functional of copula C_2 of X_i and X_j.
  - Replace C_2 in the formula by a copula C corresponding to $\boldsymbol{X}$.
  - Normalize to ensure $\kappa_d(\Pi) = 0$ and $\kappa_d(M) = 1$.

Kendall's Tau

- Bivariate version

$$\tau(X_1, X_2) = \mathrm{P}\{(X_1 - Y_1)(X_2 - Y_2) > 0\} - \mathrm{P}\{(X_1 - Y_1)(X_2 - Y_2) < 0\}$$

  where $(X_1, X_2)^\top$ and $(Y_1, Y_2)^\top$ are iid.

- Copula expression

$$\tau(C_2) = 4 \int_{[0,1]^2} C_2(\boldsymbol{u}) dC_2(\boldsymbol{u}) - 1$$

- Generalization

$$\tau(C) = \frac{1}{2^{d-1} - 1} \left\{ 2^d \int_{[0,1]^d} C(\boldsymbol{u}) dC(\boldsymbol{u}) - 1 \right\}.$$

Spearman's Rho

- Bivariate version

$$\rho(X_1, X_2) = \frac{\operatorname{cov}(F_1(X_1), F_2(X_2))}{\sqrt{\operatorname{var}(F_1(X_1))}\sqrt{\operatorname{var}(F_2(X_2))}}.$$

- Copula expression

$$\begin{aligned}\rho(C_2) &= 12 \int_{[0,1]^2} \Pi(\boldsymbol{u}) dC_2(\boldsymbol{u}) - 3 \\ &= \frac{\int_{[0,1]^2} C_2(\boldsymbol{u}) d\boldsymbol{u} - \int_{[0,1]^2} \Pi(\boldsymbol{u}) d\boldsymbol{u}}{\int_{[0,1]^2} M(\boldsymbol{u}) d\boldsymbol{u} - \int_{[0,1]^2} \Pi(\boldsymbol{u}) d\boldsymbol{u}}.\end{aligned}$$

- Generalization

$$\rho(C) = \frac{d+1}{\{2^d - (d+1)\}} \left\{ 2^d \int_{[0,1]^d} C(\boldsymbol{u}) d\boldsymbol{u} - 1 \right\}.$$

Blomqvist's Beta

- Bivariate version

$$\beta(X_1, X_2) = \mathrm{P}\{(X_1 - \tilde{x}_1)(X_2 - \tilde{x}_2) > 0\} - \mathrm{P}\{(X_1 - \tilde{x}_1)(X_2 - \tilde{x}_2) < 0\}$$

  where X_1, X_2 have medians $\tilde{x}_1, \tilde{x}_2$.

- Copula expression

$$\begin{aligned}\beta(C_2) &= 4C_2(1/2, 1/2) - 1 \\ &= \frac{C_2(1/2,1/2) - \Pi(1/2,1/2) + \overline{C}_2(1/2,1/2) - \overline{\Pi}(1/2,1/2)}{M(1/2,1/2) - \Pi(1/2,1/2) + \overline{M}(1/2,1/2) - \overline{\Pi}(1/2,1/2)}\end{aligned}$$

- Generalization

$$\beta(C) = \frac{2^{d-1}}{2^{d-1} - 1}\left\{C(1/2, \dots, 1/2) + \overline{C}(1/2, \dots, 1/2) - 2^{1-d}\right\}.$$

Clayton Copula Example

- Clayton family of copulas

$$C(\boldsymbol{u}) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-1/\theta}$$

  for $\theta > 0$.
- Copula-based generalizations
  - $\tau(C) = \frac{1}{2^{d-1}-1} \left\{ -1 + 2^d \prod_{i=0}^{d-1} \frac{1+i\theta}{2+i\theta} \right\}$
  - $\rho(C)$ needs to be calculated via its definition using numerical integration
  - $\beta(C) =$ $\frac{2^{d-1}}{2^{d-1}-1} \left\{ (d2^{\theta} - d + 1)^{-1/\theta} + \sum_{i=0}^{d} (-1)^i \binom{d}{i} (i2^{\theta} - i + 1)^{-1/\theta} - 2^{1-d} \right\}$
  - Since Clayton's Copula is symmetric in its arguments, generalization based on bivariate averages is equal to above formulas evaluated in $d = 2$.

Clayton Copula Example

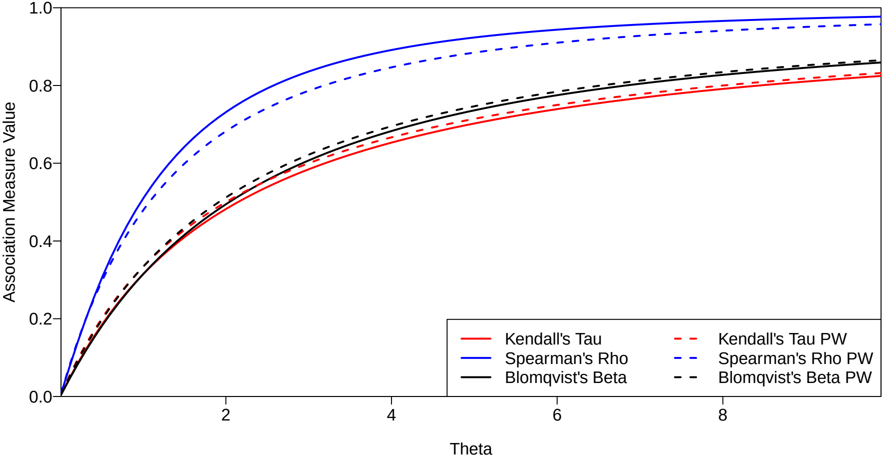

Figure: Association Measures as functions of θ for 4-dimensional Clayton Copula

Clayton Copula Example

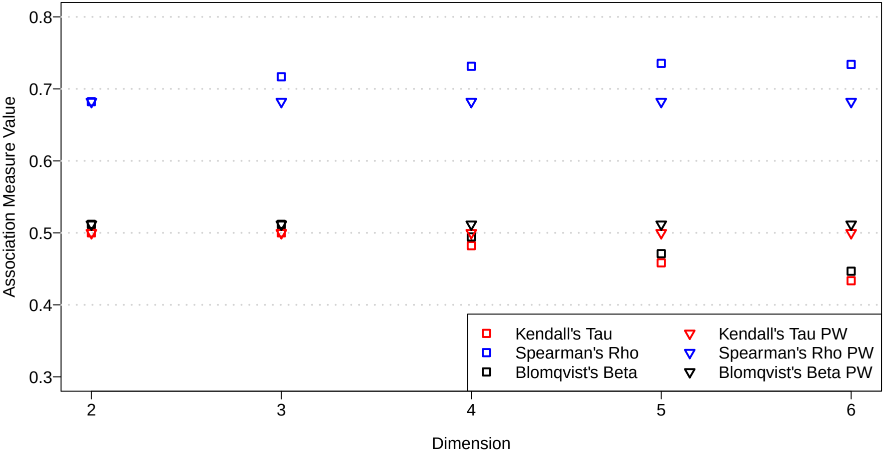

Figure: Association Measures as functions of dimension for Clayton Copula with parameter $\theta = 2$

References

F. Schmid, R. Schmidt, T. Blumentritt, S. Gaißer, and M. Ruppert.
Copula-based measures of multivariate association.
In *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, pages 209–236. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

Ch. Genest, J. Nešlehová, and N. B. Ghorbal.
Estimators based on Kendall's Tau in multivariate copula models.
Australian & New Zealand Journal of Statistics, 53(2):157–177, 2011.

Thank you for your attention!