

Watch out, spurious correlations ahead!

M. I. Ortego¹, J. J. Egozcue¹,

SUMMARY

Modeling dependence between two or more variables is a common issue in statistical applications. The Pearson correlation coefficient is often used to measure dependence, although it only captures linear dependence. Kendall's τ or Spearman's ρ , among others [4, 5] are popular alternatives for those wanting to model dependencies other than linear. The link between these coefficients and copula functions [1, 2] makes them an attractive alternative, as copulas seem a useful tool to represent dependence between variables.

When dealing with compositional data, the representation of dependence is a key issue [3]. In fact, the first efforts towards the suitable treatment of compositional data were driven by the burst of spurious (Pearson) correlation into the scene. Some researchers may expect that Kendall's τ , Spearman's ρ or the copula between parts of a composition are not spurious thus suggesting a way to circumvent log-ratio treatment of compositional data. This work is a warning against this simplification. In this work, these dependence coefficients and copula functions are used to measure the dependence between parts of a composition and its subcompositions. It is shown that if the compositional structure of data is ignored, the use of these measures of dependence also presents great shortcomings, as they are subcompositionally incoherent and, consequently, they are spurious measures of dependence between parts of a composition.

Keywords: compositional data, copula, dependence modeling

AMS Classification: 62H20, 62-07, 60E99

References

- [1] C. GENEST AND A.-C. FAVRE (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12(4), 347–368.
- [2] R. B. NELSEN (1999). *An introduction to copulas*. New York, NY, USA: Springer-Verlag.
- [3] V. PAWLOWSKY-GLAHN AND J. J. EGOZCUE AND R. TOLOSANA-DELGADO. (2015) *Modeling and analysis of compositional data*. John Wiley & Sons, Chichester UK.
- [4] M. SCARSINI (1984). On measures of concordance. *Stochastica* 8(3), 201–218.
- [5] B. SCHWEIZER AND E. F. WOLFF (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics* 9(4), 870–885.

¹Department of Civil and Environmental Engineering
Universitat Politècnica de Catalunya- BarcelonaTECH
email: ma.isabel.ortego@upc.edu