Gaussian copula regression using *R*

Guido Masarotto[1], Cristiano Varin[2][*]

1. Università degli Studi di Padova, Italy
2. Università Ca' Foscari, Venezia, Italy
[*]Contact author: sammy@unive.it

Marginal regression models for non-normal correlated responses are typically fitted by the popular generalized estimating equations approach of Liang and Zeger (1986). Despite several theoretical and practical advantages, likelihood analysis of non-normal marginal regression models is much less diffuse, see *e.g.* Diggle et al. (2002). The main reason is the difficult identification of general classes of multivariate distributions for categorical and discrete responses. Gaussian copulas provide a possible solution with a general framework for modelling dependent responses of any type (Song, 2000). Gaussian copulas combine the simplicity of interpretation in marginal modelling with the flexibility in the specification of the dependence structure. Despite this, Gaussian copula regression had still a limited use since for noncontinuous dependent responses the likelihood function requires the approximation of high-dimensional integrals. Masarotto and Varin (2010) propose an adaptation of the Geweke-Hajivassiliou-Keane importance sampling algorithm (Keane, 1994; Train 2003) to overcome the numerical difficulties of the likelihood inference.

The *R* package **mr** implements the methodology discusses in Masarotto and Varin (2010). The package allows a flexible specification of the marginals and the dependence structure. At the time of writing, the package contains methods for inference in regression models for longitudinal and clustered responses, time series, spatially correlated observations and cross-correlated studies. The functionality of the package will be illustrated on several real data examples arising in Biostatistics.

References

Diggle, P J , Heagerty, P, Liang, K-Y, Zeger, S L (2002). *Analysis of Longitudinal Data*. Second edition. Oxford University Press: Oxford.

Keane, M P (1994). A computationally practical simulation estimator for panel data. *Econometrica 62*, 95–116.

Liang, K L, Zeger, S L (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Masarotto, G, Varin, C (2010). Gaussian dependence models for non-Gaussian marginal regression. *Submitted*.

Song, P X-K (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics 27*, 305–320.

Train, K E (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge.