Nonparametric Estimation of Copulas with Discrete Outcomes

Lu Yang *[1], Edward W. Frees †[2], and Zhengjun Zhang‡[3]

[1]Amsterdam School of Economics, University of Amsterdam.
[2]Wisconsin School of Business, University of Wisconsin - Madison.
[3]Department of Statistics, University of Wisconsin - Madison.

Abstract

Multivariate discrete outcomes are common in a wide range of areas including insurance, finance, and biology. For instance, in property insurance, it is common that a policy contains multiple coverage types, e.g. building and contents coverage and motor vehicle coverage, so that the analyst may observe multiple claim frequencies, one from each coverage type. When the interplay between outcomes is significant, quantifying dependencies among interrelated variables is of great importance. In the foregoing example, quantifying dependencies among risks is critical for understanding the uncertainty of the portfolio, and thus is important for an insurer's solvency and profitability. Due to their ability to accommodate dependence flexibly, copulas are being applied increasingly.

Yet the application of copulas on discrete data is still in its infancy; one of the biggest barriers is the identifiability of copulas, calling into question model interpretations and predictions ([1]). We study the issue of identifiability in a regression context. As the marginal distributions vary with covariates, inclusion of continuous regressors provides a region of support for copula identifiability. We establish conditions under which copula regression models are identifiable for discrete outcomes.

In addition, since the properties of continuous outcomes do not carry over to discrete outcomes, specification of a copula model has been a problem. For any d dimensional variable $(Y_1, \ldots, Y_d)$ with marginal distribution functions $F_1(\cdot), \ldots, F_d(\cdot)$, when each Y_j, $j = 1, \ldots, d$ is continuous, the probability integral transform $F_j(Y_j)$ is uniformly distributed, and the unique underlying copula is actually the joint distribution of $(F_1(Y_1), \ldots, F_d(Y_d))$. Nonetheless, for a discrete outcome such as Y_1, the distribution of $F_1(Y_1)$ is generally not uniform, and the related copulas do not coincide with the joint distribution function of $(F_1(Y_1), \ldots, F_d(Y_d))$. Thus, the empirical copula estimators for continuous outcomes ([2]) cannot be applied directly to discrete outcomes. We propose a nonparametric estimator of copulas based on a local average approach to identify

*E-mail address: l.yang@uva.nl
†E-mail address: jfrees@bus.wisc.edu
‡E-mail address: zjz@stat.wisc.edu

the “hidden" dependence structure for discrete outcomes and develop its asymptotic properties. The proposed nonparametric estimator can also serve as a diagnostic tool for selecting a parametric form for copulas.

In the simulation study, we explore the performance of the proposed estimator under different scenarios and provide guidance on when the choice of copulas is important. The performance of the estimator improves as discreteness diminishes. The nonparametric estimator can exclude false copula models easily when the dependence is high and the discreteness level is low. A practical bandwidth selector is also proposed.

An empirical analysis examines a dataset from the Local Government Property Insurance Fund (LGPIF) in the state of Wisconsin. The LGPIF offers different types of coverage for local government properties, and we focus on the dependence between claims frequencies from building and contents coverage and motor vehicles coverage. We apply the nonparametric estimator on the data and use it to select a parametric copula model which best describes the underlying dependence structure. Compared with the nonparametric estimator, a Joe copula outperforms other commonly used copulas for this dataset, which indicates upper tail dependence.

Keywords: Identifiability; Copula specification; Insurance claim frequency.

References

[1] Genest, Christian and Nešlehová, Johanna (2007), “A primer on copulas for count data.” *Astin Bulletin*, vol. **37**(2), pp. 475–515.

[2] Deheuvels, Paul (1979), “La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance.” *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, vol. **65**(6), pp. 274–292.