**D-VINE COPULA MODEL FOR DEPENDENT BINARY DATA**

HUIHUI LIN (HLIN005@ODU.EDU) AND N. RAO CHAGANTY (RCHAGANTY@ODU.EDU)

DEPARTMENT OF MATHEMATICS AND STATISTICS
OLD DOMINION UNIVERSITY

**INTRODUCTION**

High-dimensional dependent binary data are prevalent in a wide range of scientific disciplines. A popular method for analyzing such data is the Multivariate Probit (MP) model [1]. But the MP model often fails even within a feasible range of binary correlations, because the underlying correlation matrix of the latent variables may not be positive definite. In this research we propose pair copula models, assuming the dependence between the binary variables is first order autoregressive (AR(1)) or equicorrelated structure.

The outline of this poster presentation is as follows. We start with the definition of the copula and pictorially illustrate the relation between the copula parameter and the binary correlation. We illustrate pair copula constructions of multivariate binary distributions using D-vines and C-vines. We show the application of our method on real life data. Finally, we briefly discuss our ongoing research.

**D-VINES AND C-VINES**

A multivariate copula is a cumulative distribution function (CDF) with uniform (0,1) univariate marginals [2]. The bivariate copula is sufficient for our pair copula model because a multivariate copula can be constructed using a D-vine and bivariate copula margins. We denote a bivariate copula by $C(u_1, u_2)$, where $\theta$ is the correlation coefficient if the copula is Gaussian, otherwise it is simply a parameter for Clayton, Gumbel and Frank copulas.

**COPULA DEFINITION**

Let $q_{10} = Pr(Y_1 = 0 | Y_2 = 0) = C_{10}(q_1, q_2)$, and $q_{11}$ similarly.

**BIVARIATE BINARY VARIABLES**

$$C(0, 0) = q_{10} = C_{10}(q_1, q_2)$$

**TRIVARIATE BINARY VARIABLES**

$$C(0, 1, 0) = q_1 - C(q_1, q_2)$$

**EX.AR(1) TRIVARIATE BINARY VAR.**

$$C(1, 1, 0) = 1 - q_1 - q_2 + C(q_1, q_2).$$

**REAL DATA EXAMPLE**


**REFERENCES**

