J. Ind. Soc. Agril. Statist.
61(2), 2007 : 179-191

Copula Functions for Modelling Dependence Structure with Applications in the Analysis of Clinical Data

Pranesh Kumar[1] and Mohamed M. Shoukri[2]
King Faisal Specialist Hospital and Research Center, MBC #03,
P. O. Box 3354, Riyadh 11211, Saudi Arabia

SUMMARY

Since the Pearson's linear correlation coefficient is not a complete and accurate description of dependence structure between variables even when there exists a straight-line relationship between them, copula as an alternative dependence measure is described. Copulas allow modelling linear and non-linear dependence using any choice of marginal distributions. Since many families of copulas are known, copula based approach provides flexibility in modelling and simulating the data. We have illustrated how to compute copula functions and use them to simulate data by considering a clinical trial of epileptic patients suffering from simple or complex partial seizures. A comparison with the correlation based analysis has indicated that the suggested copula based methodology is more appropriate and is capable of modelling the skewed behavior of measurements which correlation model fails to do.

Key words : Copula functions, Dependence measures, Archimedean copulas.

1. INTRODUCTION

Clinical data may arise primarily from a prospective, retrospective, case-control, clinical or longitudinal study. A particular study may result from a sequence of experiments, each one leading to the next. Possible studies may range from small laboratory experiments to the large and expensive experiments involving humans, to observational studies. Statistical methodologies are helpful in placing interpretations and inferences in their proper context. Appreciation of statistical methodology often leads to the design of study with increased precision and consequently a smaller sample size. Most biomedical or clinical data are multivariate (multifactor). In the multivariate situation, in addition to describing the frequency with which each value of each variable occurs, it is also of interest to study the association and relationship among the risk factors. Pearson's correlation coefficient, non parametric correlations like Kendall's and Spearman's rank correlations and multiple regressions are often applied to study the association and relationship. Embrechts *et al.* (1999) have critically examined that the Pearson's correlation and hence the methodologies based on this measure do not possess the desired properties of a good dependence measure. In particular, this measure fails to describe the tail-end (skewed) behavior of data or the extreme endpoints. In most survival and clinical studies, data distributions are fat-tailed and non-elliptical and thus the usual analyses based on the Pearson's correlation are not appropriate. There are number of ways to discuss and to measure dependence between random variables. Jogdeo (1982) states: "Dependence relation between random variables is one of the most widely studied subject in probability and statistics. The nature of the dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated."

To study the dependence structure, Sklar (1959) used copula to describe functions which join together one-dimensional distribution functions to form multivariate distribution functions. Copulas, however, are a recent

[1] *University of Northern British Columbia, Prince George, BC, Canada*
[2] *Schulich School of Medicine, University of Western Ontario, London, Ontario, Canada*