Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 2889–2906
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Sampling algorithms for generating joint uniform distributions using the vine-copula method

D. Kurowicka*, R.M. Cooke

Delft Institute for Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands

Received 11 October 2005; received in revised form 17 October 2006; accepted 4 November 2006

Abstract

An n-dimensional joint uniform distribution is defined as a distribution whose one-dimensional marginals are uniform on some interval I. This interval is taken to be [0,1] or, when more convenient $[-\frac{1}{2}, \frac{1}{2}]$. The specification of joint uniform distributions in a way which captures intuitive dependence structures and also enables sampling routines is considered. The question whether every n-dimensional correlation matrix can be realized by a joint uniform distribution remains open. It is known, however, that the *rank* correlation matrices realized by the joint normal family are sparse in the set of correlation matrices. A joint uniform distribution is obtained by specifying conditional rank correlations on a regular vine and a copula is chosen to realize the conditional bivariate distributions corresponding to the edges of the vine. In this way a distribution is sampled which corresponds exactly to the specification. The relation between conditional rank correlations on a vine and correlation matrix of corresponding distribution is complex, and depends on the copula used. Some results for the elliptical copulae are given.

Keywords: Joint uniform; Joint distribution; Dependence modeling; Vines; Sampling

1. Introduction

The problem of constructing and sampling distributions with given continuous invertible marginals and given rank correlation matrix, or equivalently, constructing and sampling joint uniforms with given correlation matrix, remains open. That is, we do not know whether an arbitrary correlation matrix can be realized by a joint uniform distribution.[1] We seek methods for specifying and sampling joint uniform distributions. Existing methods for generating joint uniform distributions appeal to the joint normal transformation (Iman and Conver, 1982) or the dependence tree-copula method (Cooke, 1997) where a *copula* is a bivariate distribution with uniform margins. For background on copulae see Genest and Rivest (1993), Nelsen (1999), Dall'Aglio et al. (1991) and Joe (1997).

Using the joint normal transform method, we start with a correlation matrix R which we would like to realize in a joint uniform distribution. We construct a joint normal distribution with correlation matrix R and then transform the one-dimensional margins to uniform. This transformation is not linear and does not preserve R. Hence, we do not realize a joint uniform with correlation matrix R (Ghosh and Henderson, 2002); indeed, we do not know if such a distribution

* Corresponding author. Tel.: +31 152785756; fax: +31 152687255.
E-mail addresses: d.kurowicka@tudelft.nl (D. Kurowicka), r.m.cooke@tudelft.nl (R.M. Cooke).

[1] While this manuscript was in press we received an article (Joe, 2006) disproving the conjection that every correlation matrix is a rank correlation matrix.

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.11.043

exists. In practice, for high dimensions, the matrix R will be only partially specified, and we face the notorious matrix completion problem: can a partially specified matrix be extended to a positive definite matrix.

The tree-copula method builds high-dimensional distributions from two-dimensional margins whose overlap structure forms a tree. The copulae used were either minimum information copulae or the diagonal band (Meeuwissen and Bedford, 1997; Cooke and Waij, 1986). The tree-copula method yields distributions which exactly correspond to the specification, but for n-dimensional problems we can specify only $n-1$ correlations corresponding to the edges of the tree.

Regular vines (Cooke, 1997) are a graphical tool for specifying conditional bivariate constraints. When these constraints are associated with partial correlations, it has been shown (Bedford and Cooke, 2002) that any assignment of values from $(-1, 1)$ to the partial correlations corresponding to edges on the vine is consistent, and determines a unique correlation matrix, and that every correlation matrix arises in this way. In other words, a partial correlation regular vine provides an algebraically independent parametrization of the set of correlation matrices. However, we do not know how to sample partial correlation vines, and we are unable to construct a joint uniform distribution from a partial correlation vine.

This paper gives algorithms for sampling regular vines when the bivariate constraints are associated with conditional rank correlations. The relation between conditional rank correlation and partial correlation is complex, and depends on the copula. Some results are given in Section 4. There are two advantages to using the vine-copula method with conditional rank correlations. First, we can construct and sample a distribution that exactly corresponds to the vine specification, and second, if some conditional correlations are unspecified, then a minimal information distribution satisfying the incomplete specification can easily be constructed whenever the copula makes uncorrelated margins (conditionally) independent. In this case it is simply a matter of assigning conditional rank correlation zero to the unspecified nodes in the vine (Cooke, 1997; Bedford and Cooke, 2002). The relation of incomplete vine specification to the matrix completion problem is studied in Kurowicka and Cooke (2003).

The second section reviews briefly facts about rank, product moment and partial correlations, copulas and vines. We show that an arbitrary correlation matrix need not be the rank correlation matrix of a joint normal distribution. The third section presents a sampling algorithm to exactly sample a high-dimensional distribution with uniform margins and given conditional rank correlations using the vine-copula method. The fourth section derives results concerning the relationship between conditional rank and partial correlations for copulae. The fifth section presents simulation results for the elliptical, diagonal band and Frank's copulae. The last section gives conclusions.

2. Correlation and vines

2.1. Rank, product moment, and partial correlations

The obvious relationship between product moment and rank correlations follows directly from their definitions as rank correlation is just a product moment correlation of variables transformed to uniforms. Hence for uniform variables rank and product moment correlations are equal but in general they are different.

Pearson (1907), proved that if vector (X_1, X_2) has a joint normal distribution, then the relationship between rank (r) and product moment correlation (ρ) is given by

$$\rho(X_1, X_2) = 2 \sin\left(\frac{\pi}{6} r(X_1, X_2)\right). \tag{1}$$

The proof of this fact is based on the property that the derivative of the density function for bivariate normals with respect to correlation is equal to the second order derivative with respect to x_1 and x_2.

The rank correlation has some important advantages over the product moment correlation. It always exists, can take any value in the interval $[-1, 1]$, is independent of the marginal distributions and is invariant under monotone transformations.

In this paper we will study relationships between conditional product moment, conditional rank and partial correlations. We assume that the reader is familiar with first two correlation coefficients and we introduce here only the partial correlation.