Bivariate Kumaraswamy models involving use of Arnold-Ng copulas

Barry C. Arnold

Department of Statistics, University of California, Riverside, USA

Indranil Ghosh

Department of Mathematics and Statistics, University of North Carolina, Wilmington, USA

Abstract

Bivariate Kumaraswamy distributions and several related variants have received a significant amount of interest over the last few years. However, construction of bivariate and multivariate Kumaraswamy distributions using copulas has not been explored in detail. In this article, we discuss and explore various copula based methods to construct bivariate Kumaraswamy distributions. In particular, detailed discussion of bivariate copulas based on Arnold-Ng(2011) approach is provided. Extensions to trivariate and multivariate setting are discussed. For illustrative purposes, a Brazilian HIV data set is utilized to exhibit the utility of two nested models in the class of bivariate Kumaraswamy models based on Arnold-Ng (2011) copulas.

Key Words: Bivariate Kumaraswamy distribution, Copula based construction, Arnold-Ng copula

1 Introduction

Over the last decade or so, there has been a growing interest in constructing various bivariate distributions and studying their dependence structure. For an excellent survey of these constructions the interested reader can refer to Balakrishnan et al.(2009)

and the references therein. Of late, copula based methods of construction have received a considerable amount of attention, mainly due to their analytical tractability in the context of discussing dependence structure between the coordinate random variables. A copula is a multivariate distribution function whose marginals are uniform. It couples or links the marginal distribution to their joint distribution. In order to obtain a bivariate/multivariate distribution function, one needs to simply combine two (in the bivariate case) and/or several marginal distribution functions with a selected copula function. Consequently, for the purpose of statistical modeling, it is desirable to have a broad spectrum of copulas at one's disposal.

Kumaraswamy (1980) introduced a two parameter absolutely continuous distribution originally to be used for double bounded random processes with hydrological applications. The Kumaraswamy pdf is unimodal, uniantimodal, increasing, decreasing or constant depending (similar to the beta distribution) on the values of the parameters. However, the construction of bivariate Kumaraswamy distributions has received limited attention. Barreto-Souza and Lemonte (2013) introduced a bivariate Kumaraswamy distribution related to a Marshall- Olkin survival copula and discussed some structural properties of this bivariate Kumaraswamy distribution. Arnold and Ghosh (2016 a) discussed some different strategies for constructing legitimate bivariate Kumaraswamy models via conditional specification, conditional survival function specification, and via an Arnold-Ng bivariate beta construction approach. In another paper, Arnold and Ghosh (2016 b) discussed several other construction strategies, in particular focusing on the Arnold-Ng bivariate beta distribution approach and the Kotz-Nadarajah approach, but also considering several other approaches to the problem of the constructing bivariate beta and Kumaraswamy distributions. Very recently, Ghosh and Ray (2016) discussed some copula based approaches to the construction of several bivariate Kumaraswamy type models together with an application to a real life data set focusing on financial risk assessment.

In this paper, we develop a spectrum of bivariate Kumaraswamy distributions by utilizing certain multi-parameter families of copulas that are special cases of the 8-parameter bivariate beta distribution introduced in Arnold and Ng (2011). The array of copulas introduced here can be expected to be useful for constructing other bivariate models, besides the Kumaraswamy case, after applying appropriate marginal transformations. Because of the added flexibility evident in these multiparameter (of dimension 1,2,3 or 4) copulas, they can be expected to outperform simpler copula models such as the one parameter Farlie-Gumbel-Morgenstern model. Indeed, we expect that the copula sections (i.e., Sections 2 and 3) will be of interest to a wider audience than the sections dealing specifically with the bivariate Kumaraswamy case.

2 Arnold-Ng copulas

We begin by considering the Arnold-Ng 8-parameter bivariate beta of the second kind model as follows:

$$(V_1, V_2) = \left(\frac{U_1 + U_5 + U_7}{U_3 + U_6 + U_8}, \frac{U_2 + U_5 + U_8}{U_4 + U_6 + U_7} \right), \tag{2.1}$$

where the U_i's are independent gamma distributed random variables with $U_i \sim \Gamma(\alpha_i, 1)$, $i = 1, 2, ..., 8$. This model was introduced in Arnold and Ng (2011). This is readily transformed to yield a corresponding 8-parameter bivariate beta of the first (or usual) kind model by defining

$$(Y_1, Y_2) = \left(\frac{V_1}{1 + V_1}, \frac{V_2}{1 + V_2}. \right) \tag{2.2}$$

In order to have a copula model, i.e., a model of the form (2.2), with $Uniform(0, 1)$ marginals, we must impose the following 4 constraints.

$$\alpha_1 + \alpha_5 + \alpha_7 = 1, \tag{2.3}$$

$$\alpha_2 + \alpha_5 + \alpha_8 = 1, \tag{2.4}$$

$$\alpha_3 + \alpha_6 + \alpha_8 = 1, \tag{2.5}$$

$$\alpha_4 + \alpha_6 + \alpha_7 = 1. \tag{2.6}$$

In addition, recall that all α_i's are non-negative. The resulting four dimensional parameter space may be described as follows:

$$\alpha_5 \in [0, 1], \tag{2.7}$$

$$\alpha_6 \in [0, 1], \tag{2.8}$$

$$0 \leq \alpha_7 \leq 1 - \max\{\alpha_5, \alpha_6\}, \tag{2.9}$$

$$0 \leq \alpha_8 \leq 1 - \max\{\alpha_5, \alpha_6\}. \tag{2.10}$$

The remaining α_i's, $i = 1, 2, 3, 4$, are then determined by equations (2.3)-(2.6).
Such models will be referred to as Arnold-Ng (henceforth AN) copulas. A two parameter submodel of possible interest is obtained by setting $\alpha_7 = \alpha_8 = 0$. The model is then parameterized by $(\alpha_5, \alpha_6) \in (0, 1)^2$ with the remaining α's determined by:

$$\alpha_1 = \alpha_2 = 1 - \alpha_5, \tag{2.11}$$

$$\alpha_3 = \alpha_4 = 1 - \alpha_6. \tag{2.12}$$

Five other similar two parameter families of copulas correspond to cases in which other sets of two of the parameters $\alpha_5, \alpha_6, \alpha_7, \alpha_8$ are set equal to 0.

The remainder of the article is organized in the following way. In section 3, we discuss a comprehensive list of distinct copula based models that can be derived from the general multiparameter AN copula based model, by setting some of the α_i's equal to 0.. In section 4, a brief discussion is provided regarding the construction of a bivariate Kumaraswamy distributions via such copulas. Section 5 deals with several other popular bivariate copula models which can be used to construct bivariate Kumaraswamy distributions. In section 6, we discuss some trivariate Kumaraswamy models based on gamma based trivariate copulas. In section 7, a Brazilian HIV data set is used as an illustrative application for comparing two nested bivariate Kumaraswamy models along with other copula based bivariate Kumaraswamy models. Some concluding remarks are provided in section 8.

3 A catalog of submodels

For completeness we will list all of the 17 distinct one parameter models that are included in (2.1) and can thus be transformed into copulas using (2.2).

- $\left(\frac{U_5}{U_3+U_6}, \frac{U_5}{U_4+U_6}\right)$ where $\alpha_3 = 1-\alpha, \alpha_4 = 1-\alpha, \alpha_5 = 1, \alpha_6 = \alpha$.
- $\left(\frac{U_1+U_5}{U_6}, \frac{U_2+U_5}{U_6}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\alpha, \alpha_5 = \alpha, \alpha_6 = 1$.
- $\left(\frac{U_7}{U_3+U_8}, \frac{U_2+U_8}{U_7}\right)$ where $\alpha_2 = 1-\alpha, \alpha_3 = 1-\alpha, \alpha_7 = 1, \alpha_8 = \alpha$.
- $\left(\frac{U_1+U_7}{U_8}, \frac{U_8}{U_4+U_7}\right)$ where $\alpha_1 = 1-\alpha, \alpha_4 = 1-\alpha, \alpha_7 = \alpha, \alpha_8 = 1$.
- $\left(\frac{U_5+U_7}{U_6+U_8}, \frac{U_5+U_8}{U_6+U_7}\right)$ where $\alpha_5 = 1-\alpha, \alpha_6 = 1-\alpha, \alpha_7 = \alpha, \alpha_8 = \alpha$.
- $\left(\frac{U_1+U_5}{U_3}, \frac{U_2+U_5}{U_4}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\alpha, \alpha_3 = 1, \alpha_4 = 1, \alpha_5 = \alpha$
- $\left(\frac{U_1}{U_3+U_6}, \frac{U_2}{U_4+U_6}\right)$ where $\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 1-\alpha, \alpha_4 = 1-\alpha, \alpha_6 = \alpha$

- $\left(\frac{U_1+U_7}{U_3}, \frac{U_2}{U_4+U_7}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1, \alpha_3 = 1, \alpha_4 = 1-\alpha, \alpha_7 = \alpha$
- $\left(\frac{U_1}{U_3+U_8}, \frac{U_2+U_8}{U_4}\right)$ where $\alpha_1 = 1, \alpha_2 = 1-\alpha, \alpha_3 = 1-\alpha, \alpha_4 = 1, \alpha_8 = \alpha$
- $\left(\frac{U_1+U_7}{U_3+U_6}, \frac{U_2}{U_6+U_7}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1, \alpha_3 = \alpha, \alpha_6 = 1-\alpha, \alpha_8 = \alpha$
- $\left(\frac{U_1}{U_6+U_8}, \frac{U_2+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1, \alpha_2 = 1-\alpha, \alpha_4 = \alpha, \alpha_6 = 1-\alpha, \alpha_8 = \alpha$
- $\left(\frac{U_1+U_7}{U_6+U_8}, \frac{U_2+U_8}{U_6+U_7}\right)$ where $\alpha_1 = \alpha, \alpha_2 = \alpha, \alpha_6 = \alpha, \alpha_7 = 1-\alpha, \alpha_8 = 1-\alpha$
- $\left(\frac{U_1+U_5}{U_3+U_8}, \frac{U_5+U_8}{U_4}\right)$ where $\alpha_1 = 1-\alpha, \alpha_3 = \alpha, \alpha_4 = 1, \alpha_5 = \alpha, \alpha_8 = 1-\alpha$
- $\left(\frac{U_1+U_5}{U_6+U_8}, \frac{U_5+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1-\alpha, \alpha_4 = 1-\alpha, \alpha_5 = \alpha, \alpha_6 = \alpha, \alpha_8 = 1-\alpha$
- $\left(\frac{U_5+U_7}{U_3}, \frac{U_2+U_5}{U_4+U_7}\right)$ where $\alpha_2 = \alpha, \alpha_3 = 1, \alpha_4 = 1-\alpha, \alpha_5 = 1-\alpha, \alpha_7 = \alpha$
- $\left(\frac{U_5+U_7}{U_3+U_6}, \frac{U_2+U_5}{U_6+U_7}\right)$ where $\alpha_2 = \alpha, \alpha_3 = \alpha, \alpha_5 = 1-\alpha, \alpha_6 = 1-\alpha, \alpha_7 = \alpha$
- $\left(\frac{U_5+U_7}{U_3+U_8}, \frac{U_5+U_8}{U_4+U_7}\right)$ where $\alpha_3 = 1-\alpha, \alpha_4 = 1-\alpha, \alpha_5 = 1-\alpha, \alpha_7 = \alpha, \alpha_8 = \alpha$

In addition there are 18 distinct two parameter models included in (2.1) that can yield copulas via (2.2), 8 three parameter models and one four parameter model (the full model). The two parameter models are:

- $\left(\frac{U_1+U_5}{U_3+U_6}, \frac{U_2+U_5}{U_4+U_6}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\alpha, \alpha_3 = 1-\beta, \alpha_4 = 1-\beta, \alpha_5 = \alpha, \alpha_6 = \beta$
- $\left(\frac{U_1+U_5+U_7}{U_3}, \frac{U_2+U_5}{U_4+U_7}\right)$ where $\alpha_1 = 1-\alpha-\beta, \alpha_2 = 1-\alpha, \alpha_3 = 1, \alpha_4 = 1-\beta,$
  $\alpha_5 = \alpha, \alpha_7 = \beta$
- $\left(\frac{U_1+U_5}{U_3+U_8}, \frac{U_2+U_5+U_8}{U_4}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\alpha-\beta, \alpha_3 = 1-\beta, \alpha_4 = 1,$
  $\alpha_5 = \alpha, \alpha_8 = \beta$
- $\left(\frac{U_1+U_7}{U_3+U_6}, \frac{U_2}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1-\beta, \alpha_2 = 1, \alpha_3 = 1-\alpha, \alpha_4 = 1-\alpha-\beta,$
  $\alpha_6 = \alpha, \alpha_7 = \beta$

- $\left(\frac{U_1}{U_3+U_6+U_8}, \frac{U_2+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1, \alpha_2 = 1-\beta, \alpha_3 = 1-\alpha-\beta, \alpha_4 = 1-\alpha,$ $\alpha_6 = \alpha, \alpha_8 = \beta$
- $\left(\frac{U_1+U_7}{U_3+U_8}, \frac{U_2+U_8}{U_4+U_7}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\beta, \alpha_3 = 1-\beta, \alpha_4 = 1-\alpha, \alpha_7 = \alpha, \alpha_8 = \beta$
- $\left(\frac{U_1+U_5+U_7}{U_3+U_6}, \frac{U_2+U_5}{U_6+U_7}\right)$ where $\alpha_1 = \beta-\alpha, \alpha_2 = 1-\alpha, \alpha_3 = 1-\beta, \alpha_5 = \alpha, \alpha_6 = \beta,$ $\alpha_7 = 1-\beta$
- $\left(\frac{U_1+U_7}{U_3+U_6+U_8}, \frac{U_2+U_8}{U_6+U_7}\right)$ where $\alpha_1 = 1-\alpha, \alpha_2 = 1-\beta, \alpha_3 = \alpha-\beta, \alpha_6 = 1-\alpha,$ $\alpha_7 = \alpha, \alpha_8 = \beta$
- $\left(\frac{U_1+U_5}{U_6+U_8}, \frac{U_2+U_5+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1-\beta, \alpha_2 = \alpha-\beta, \alpha_4 = 1-\alpha, \alpha_5 = \beta, \alpha_6 = \alpha,$ $\alpha_8 = 1-\alpha$
- $\left(\frac{U_1+U_7}{U_6+U_8}, \frac{U_2+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1-\beta, \alpha_2 = 1-\alpha, \alpha_4 = \alpha-\beta, \alpha_6 = 1-\alpha,$ $\alpha_7 = \beta, \alpha_8 = \alpha$
- $\left(\frac{U_1+U_5+U_7}{U_6+U_8}, \frac{U_2+U_5+U_8}{U_6+U_7}\right)$ where $\alpha_1 = \alpha-\beta, \alpha_2 = \alpha-\beta, \alpha_5 = \beta, \alpha_6 = \alpha, \alpha_7 = 1-\alpha,$ $\alpha_8 = 1-\alpha$
- $\left(\frac{U_1+U_5}{U_3+U_6+U_8}, \frac{U_5+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1-\alpha, \alpha_3 = \alpha-\beta, \alpha_4 = 1-\beta, \alpha_5 = \alpha, \alpha_6 = \beta,$ $\alpha_8 = 1-\alpha$
- $\left(\frac{U_1+U_5+U_7}{U_3+U_8}, \frac{U_5+U_8}{U_4+U_7}\right)$ where $\alpha_1 = \alpha-\beta, \alpha_3 = 1-\alpha, \alpha_4 = 1-\beta, \alpha_5 = 1-\alpha,$ $\alpha_7 = \beta, \alpha_8 = \alpha$
- $\left(\frac{U_1+U_5+U_7}{U_6+U_8}, \frac{U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = \alpha-\beta, \alpha_4 = \alpha-\beta, \alpha_5 = 1-\alpha, \alpha_6 = 1-\alpha,$ $\alpha_7 = \beta, \alpha_8 = \alpha$
- $\left(\frac{U_5+U_7}{U_3+U_6}, \frac{U_2+U_5}{U_4+U_6+U_7}\right)$ where $\alpha_2 = 1-\alpha, \alpha_3 = 1-\beta, \alpha_4 = \alpha-\beta, \alpha_5 = \alpha, \alpha_6 = \beta,$ $\alpha_7 = 1-\alpha$

- $\left(\frac{U_5+U_7}{U_3+U_8}, \frac{U_2+U_5+U_8}{U_4+U_7}\right)$ where $\alpha_2 = \beta - \alpha, \alpha_3 = 1 - \alpha, \alpha_4 = 1 - \alpha, \alpha_5 = 1 - \beta,$ $\alpha_7 = \beta, \alpha_8 = \alpha$

- $\left(\frac{U_5+U_7}{U_3+U_6+U_8}, \frac{U_2+U_5+U_8}{U_6+U_7}\right)$ where $\alpha_2 = \alpha - \beta, \alpha_3 = \alpha - \beta, \alpha_5 = 1 - \alpha, \alpha_6 = 1 - \alpha,$ $\alpha_7 = \alpha, \alpha_8 \beta$

- $\left(\frac{U_5+U_7}{U_3+U_6+U_8}, \frac{U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_3 = \alpha - \beta, \alpha_4 = \alpha - \beta, \alpha_5 = \alpha, \alpha_6 = \beta, \alpha_7 = 1 - \alpha,$ $\alpha_8 = 1 - \alpha.$

The 8 three parameter models are:

- $\left(\frac{U_1+U_5+U_7}{U_3+U_6}, \frac{U_2+U_5}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1 - \alpha - \gamma, \alpha_2 = 1 - \alpha, \alpha_3 = 1 - \beta,$ $\alpha_4 = 1 - \beta - \gamma, \alpha_5 = \alpha, \alpha_6 = \beta, \alpha_7 = \gamma$

- $\left(\frac{U_1+U_5}{U_3+U_6+U_8}, \frac{U_2+U_5+U_8}{U_4+U_6}\right)$ where $\alpha_1 = 1 - \alpha, \alpha_2 = 1 - \alpha - \gamma, \alpha_3 = 1 - \beta - \gamma,$ $\alpha_4 = 1 - \beta, \alpha_5 = \alpha, \alpha_6 = \beta, \alpha_8 = \gamma$

- $\left(\frac{U_1+U_5+U_7}{U_3+U_8}, \frac{U_2+U_5+U_8}{U_4+U_7}\right)$ where $\alpha_1 = 1 - \alpha - \gamma, \alpha_2 = 1 - \beta - \gamma, \alpha_3 = 1 - \beta,$ $\alpha_4 = 1 - \alpha, \alpha_5 = \gamma, \alpha_7 = \alpha, \alpha_8 = \beta$

- $\left(\frac{U_1+U_7}{U_3+U_6+U_8}, \frac{U_2+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1 - \alpha, \alpha_2 = 1 - \beta, \alpha_3 = 1 - \beta - \gamma,$ $\alpha_4 = 1 - \alpha - \gamma, \alpha_6 = \gamma, \alpha_7 = \alpha, \alpha_8 = \beta$

- $\left(\frac{U_1+U_5+U_7}{U_3+U_6+U_8}, \frac{U_2+U_5+U_8}{U_6+U_7}\right)$ where $\alpha_1 = 1 - \alpha - \gamma, \alpha_2 = 1 - \beta - \gamma, \alpha_3 = \alpha - \beta, \alpha_5 = \gamma,$ $\alpha_6 = 1 - \alpha, \alpha_7 = \alpha, \alpha_8 = \beta$

- $\left(\frac{U_1+U_5+U_7}{U_6+U_8}, \frac{U_2+U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1 - \beta - \gamma, \alpha_2 = 1 - \alpha - \gamma, \alpha_4 = \alpha - \beta, \alpha_5 = \gamma,$ $\alpha_6 = 1 - \alpha, \alpha_7 = \beta, \alpha_8 = \alpha$

- $\left(\frac{U_1+U_5+U_7}{U_3+U_6+U_8}, \frac{U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = \alpha - \beta, \alpha_3 = 1 - \alpha - \gamma, \alpha_4 = 1 - \beta - \gamma,$ $\alpha_5 = 1 - \alpha, \alpha_6 = \gamma, \alpha_7 = \beta, \alpha_8 = \alpha$

- $\left(\frac{U_5+U_7}{U_3+U_6+U_8}, \frac{U_2+U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_2 = \beta - \alpha, \alpha_3 = 1 - \alpha - \gamma, \alpha_4 = 1 - \beta - \gamma,$ $\alpha_5 = 1 - \beta, \alpha_6 = \gamma, \alpha_7 = \beta, \alpha_8 = \alpha.$

Finally we restate the full model parameterization as follows.

$\left(\frac{U_1+U_5+U_7}{U_3+U_6+U_8}, \frac{U_2+U_5+U_8}{U_4+U_6+U_7}\right)$ where $\alpha_1 = 1 - \beta - \delta, \alpha_2 = 1 - \alpha - \delta, \alpha_3 = 1 - \alpha - \gamma,$ $\alpha_4 = 1 - \beta - \gamma, \alpha_5 = \delta, \alpha_6 = \gamma, \alpha_7 = \beta, \alpha_8 = \alpha$

There are obvious parametric constraints on the parameter spaces of these models, determined by the requirement that all of the α_i's must be in the interval $[0, 1]$.

4 Bivariate Kumaraswamy models using Arnold-Ng copulas

The Kumaraswamy distribution on the interval $(0, 1)$, has its probability density function (pdf) and its cdf with two shape parameters $\gamma > 0$ and $\delta > 0$ defined by

$$f(x) = \gamma\delta x^{\gamma-1}(1 - x^{\gamma})^{\delta-1} \quad \text{and } F(x) = 1 - (1 - x^{\gamma})^{\delta}. \tag{4.1}$$

The corresponding quantile function has the form

$$F^{-1}(u) = \left[1 - (1 - u)^{1/\delta}\right]^{1/\gamma}, \quad 0 < u < 1. \tag{4.2}$$

Transforming the copula (2.2) marginally with quantile functions of the form (4.2), we obtain a bivariate Kumaraswamy model of the following form

$$(X_1, X_2) = (\left[1 - (1 - Y_1)^{1/\delta}\right]^{1/\gamma}, \left[1 - (1 - Y_2)^{1/\delta}\right]^{1/\gamma}), \tag{4.3}$$

where Y_1, Y_2 are as in (2.2) and are functions of the independent gamma distributed U_i's.

Note that, in general, the model (4.3) will not have a density function available in closed form. It is also worth remarking that the model (without any simplifying parametric constraints on the α's) is an 8 parameter model. This is the same as the dimension of the parameter space of the Kumaraswamy- conditionals distribution.

If instead, we marginally transform any of the copulas corresponding to the submodels listed in Section 3, we will have developed bivariate Kumaraswamy distributions with more restricted parameter spaces. In some very simple cases, a joint density for the resulting distribution will be available.

To illustrate the use of these Arnold-Ng-Bivariate-Kumaraswamy models, we will consider two simple submodels obtained by setting in (2.1), in one case, $\alpha_5 = \alpha_6 = 0, \alpha_7 = \alpha$ and $\alpha_8 = \beta$, and in the other case, the submodel obtained by setting $\alpha_5 = \alpha_6 = 0, \alpha_7 = \alpha_8 = \alpha$.. The corresponding two parameter family of copulas is of the form

$$(Y_1, Y_2) = \left(\frac{U_1 + U_7}{U_1 + U_7 + U_3 + U_8}, \frac{U_2 + U_8}{U_2 + U_8 + U_4 + U_7} \right), \tag{4.4}$$

where

$$\alpha_1 = 1 - \alpha, \quad \alpha_2 = 1 - \beta, \quad \alpha_3 = 1 - \beta, \tag{4.5}$$

$$\alpha_4 = 1 - \alpha, \quad \alpha_7 = \alpha, \quad \alpha_8 = \beta. \tag{4.6}$$

Here $(\alpha, \beta) \in (0, 1)^2$. The one parameter submodel is obtained by setting $\beta = \alpha$, i.e.,

$$\alpha_1 = 1 - \alpha, \quad \alpha_2 = 1 - \alpha, \quad \alpha_3 = 1 - \alpha, \tag{4.7}$$

$$\alpha_4 = 1 - \alpha, \quad \alpha_7 = \alpha, \quad \alpha_8 = \alpha. \tag{4.8}$$

Marginal transformations of the form (4.3) are then applied to yield two nested bivariate Kumaraswamy models with 6 and 5 parameters respectively.

It is to be expected that the 6 parameter model will fit data better than the restricted

5 parameter model. Some competing copula based bivariate Kumaraswamy models will be reviewed in the next section.

5 Bivariate Kumaraswamy models via other copulas

Instead of utilizing the gamma based copulas described in Sections 2 and 3, one might consider use of one of the copulas more frequently used in the literature. For example, one might consider use of a standard Farlie-Gumbel-Morgenstern copula of the form

$$F(u_1,u_2) = u_1u_2[1+\theta(1-u_1)(1-u_2)], \quad (u_1,u_2) \in (0,1)^2, \tag{5.1}$$

where $\theta \in [-1,1]$. After marginal transformations of the form (4.3) this yields a 5 parameter bivariate Kumaraswamy model. The advantage possessed by this FGM-K model is that analytic expressions are available for its joint distribution and density functions.

Other bivariate Kumaraswamy models based on popular competing copulas can be developed using the following approach and notation.

Suppose that (U_1,U_2) has joint distribution $C(u_1,u_2)$, a copula with uniform marginals. Now define $(X_1,X_2) = (F_{X_1}^{-1}(U_1), F_{X_2}^{-1}(U_2))$. It is then readily verified that

$$F_{X_1,X_2}(x_1,x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$

and that the marginal distributions of X_1 and X_2 are F_{X_1} and F_{X_2} respectively. For example, we can use the Ali-Mikhail-Haq copula

$$C_{AMH}(u_1,u_2) = \frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)} \tag{5.2}$$

to obtain the following bivariate AMH Kumarawamy distribution with Kumaraswamy marginals,

$$F_{X_1,X_2}(x_1,x_2) = \frac{[1-(1-x_1^a)^b][1-(1-x_2^c)^d]}{1-\theta(1-x_1^a)^b(1-x_2^c)^d}.$$

This is a five parameter model. Another popular copula is the normal copula defined as follows

$$C_{Normal}(u_1,u_2) = \Phi^{(2)}(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho),$$

where $\Phi^{(2)}(x_1,x_2;\rho)$ is the joint distribution function of a bivariate normal vector with zero means, unit variances and correlation ρ. The corresponding bivariate normal copula distribution with Kumaraswamy marginals is then

$$F_{X_1,X_2}(x_1,x_2) = \Phi^{(2)}(\Phi^{-1}(1-(1-x_1^a)^b), \Phi^{-1}(1-(1-x_2^c)^d); \rho).$$

This too is a 5 parameter model. Yet another popular choice is the Gumbel-Hougard copula

$$C_{G-H}(u_1,u_2) = \exp\{-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{1/\alpha}\} \tag{5.3}$$

yielding the following 5 parameter bivariate Kumaraswamy distribution

$$F_{X_1,X_2}(x_1,x_2) = \exp\{-[(-\ln(1-(1-x_1^a)^b))^\alpha + (-\ln(1-(1-x_2^c)^d))^\alpha]^{1/\alpha}\}.$$

There are two possible Marshall- Olkin copulas that can be utilized. Both begin by considering the Marshall- Olkin exponential model with distribution function

$$F_{X_1,X_2}(x_1,x_2) = 1 - e^{-(1+\delta)x_1} - e^{-(1+\delta)x_2} + e^{-x_1}e^{-x_2}e^{-\delta\max\{x_1,x_2\}},$$

and with the more recognizable survival function

$$\overline{F}_{X_1,X_2}(x_1,x_2) = e^{-x_1}e^{-x_2}e^{-\delta\max\{x_1,x_2\}}.$$

Note that $P(X_1 > x_1) = e^{-(1+\delta)x_1}$ and $P(X_2 > x_2) = e^{-(1+\delta)x_2}$ For the first copula, we define

$$(U_1, U_2) = (1 - e^{-(1+\delta)X_1}, 1 - e^{-(1+\delta)X_2}).$$

Clearly U_1 and U_2 are uniform(0,1) variables, so the joint distribution of (U_1, U_2) will be a copula. In this case we have

$$\begin{aligned} C^{(1)}_{M-O}(u_1, u_2) &= P(U_1 \leq u_1, U_2 \leq u_2) \\ &= P(1 - e^{-(1+\delta)X_1} \leq u_1, 1 - e^{-(1+\delta)X_2} \leq u_2) \\ &= P(X_1 \leq -\ln[(1-u_1)^{1/(1+\delta)}], X_2 \leq -\ln[(1-u_2)^{1/(1+\delta)}] \\ &= 1 - (1-u_1) - (1-u_2) + (1-u_1)^{1/(1+\delta)}(1-u_2)^{1/(1+\delta)}(1 - \min\{u_1, u_2\})^{\delta/(1+\delta)}. \end{aligned}$$

A more attractive Marshall-Olkin copula is constructed by defining

$$(U_1, U_2) = (e^{-(1+\delta)X_1}, e^{-(1+\delta)X_2}).$$

Clearly, here too, U_1 and U_2 are uniform (0,1) variables, so their joint distribution will be a copula. In this case we have

$$C^{(2)}_{M-O}(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2) = P(e^{-(1+\delta)X_1} \leq u_1, e^{-(1+\delta)X_2} \leq u_2)$$

$$= P(X_1 > -\ln[u_1^{1/(1+\delta)}], X_2 > -\ln[u_2^{1/(1+\delta)}]$$

$$= u_1^{1/(1+\delta)} u_2^{1/(1+\delta)} [\min\{u_1, u_2\}]^{\delta/(1+\delta)},$$

leading to the bivariate Kumaraswamy model

$$F_{X_1,X_2}(x_1,x_2) = (1-(1-x_1^a)^b)^{1/(1+\delta)}(1-(1-x_2^c)^d)^{1/(1+\delta)}(\min\{(1-(1-x_1^a)^b),(1-(1-x_1^c)^d)\})^{\delta/(1+\delta)}. \tag{5.4}$$

Note that this model is distinct from the Barreto-Souza-Lemonte (2013) model. The B-S-L model has joint survival given by

$$P(X_1 > x_1, X_2 > x_2) = (1-x_1^a)^{b_1}(1-x_2^a)^{b_2}(1-(\max\{x_1,x_2\})^a)^{b_3}$$

with marginals $P(X_1 > x_1) = (1-x_1^a)^{b_1+b_3}$ and $P(X_2 > x_2) = (1-x_2^a)^{b_2+b_3}$.

A copula can also be used to directly define a bivariate survival function. If it is so used, it is called a survival copula. We will outline this approach here, but we will not provide a detailed discussion. If we use the notation

$$\overline{F}_{X_1,X_2}(x_1,x_2) = P(X_1 > x_1, X_2 > x_2)$$

then we can use a copula C to define a joint survival function by

$$\overline{F}_{X_1,X_2}(x_1,x_2) = C(\overline{F}_{X_1}(x_1), \overline{F}_{X_2}(x_2))$$

and this joint distribution will have marginal survival functions $\overline{F}_{X_1}(x_1)$ and $\overline{F}_{X_2}(x_2)$. These survival copula models are distinct from the corresponding copula models. For example a bivariate AMH Kumarawamy distribution with Kumaraswamy marginals constructed via the survival copula approach will be of the form

$$\overline{F}_{X_1,X_2}(x_1,x_2) = \frac{(1-x_1^a)^b(1-x_2^c)^d}{1-\theta[1-(1-x_1^a)^b][1-(1-x_2^c)^d]},$$

and the second Marshall Olkin distribution will yield:

$$\overline{F}_{X_1,X_2}(x_1,x_2) = (1-x_1^a)^{b/(1+\delta)}(1-x_2^c)^{d/(1+\delta)}[\min\{(1-x_1^a)^b),(1-x_2^c)^d\}]^{\delta/(1+\delta)},$$

which is closely related to the Barreto-Souza and Lemonte (2013) model, but distinct.

A comparative study of bivariate Kumaraswamy models constructed using many of the copulas described in this sections will be the subject of a separate report.

6 On higher dimensional gamma based copulas

The three dimensional version of this construction will involve 26 U_i's. This number can be verified by noting that a trivariate model (V_1, V_2, V_3) expressed as ratios of independent sums of independent gamma variables (with unit scale parameter), will involve 6 places where a particular U can appear, three numerators and three denominators. But a particular U cannot appear in both the numerator and denominator of any of the three V_i's. There will be 6 U's which appear in one of the 6 possible places. These will be denoted by $U_1, U_2, ..., U_6$. There will be 12 U's that appear in exactly two of the 6 possible positions, denoted by $U_7, U_8, ..., U_{18}$. Finally there are 8 U's that appear in 3 places, namely $U_{19}, U_{20}, ..., U_{26}$. No U can appear in more than 3 places without violating the requirement that numerators must be independent of their corresponding denominators.

Thus, there are a total of 26 parameters in the model where $U_i,\ i = 1, 2, ..., 26$ are independent variables with $U_i \sim \Gamma(\alpha_i, 1)$ for each i. The model can then be expressed in the following, somewhat daunting, form.

$$V_1 = \frac{U_1 + U_7 + U_8 + U_9 + U_{10} + U_{19} + U_{20} + U_{21} + U_{22}}{U_4 + U_{11} + U_{12} + U_{13} + U_{14} + U_{23} + U_{24} + U_{25} + U_{26}}, \tag{6.1}$$

$$V_2 = \frac{U_2 + U_7 + U_{11} + U_{15} + U_{16} + U_{19} + U_{20} + U_{23} + U_{24}}{U_5 + U_9 + U_{13} + U_{17} + U_{18} + U_{21} + U_{22} + U_{25} + U_{26}}, \tag{6.2}$$

and

$$V_3 = \frac{U_3 + U_8 + U_{12} + U_{15} + U_{17} + U_{19} + U_{21} + U_{23} + U_{25}}{U_6 + U_{10} + U_{14} + U_{16} + U_{18} + U_{20} + U_{22} + U_{24} + U_{26}}. \tag{6.3}$$

The pattern for the dimensions of parameter spaces of the multivariate models becomes clear. The univariate model involves 2 U's, i.e., $3^1 - 1$. The bivariate model involves 8 U's, i.e., $3^2 - 1$. The trivariate case involves 26 U's, i.e., $3^3 - 1$., and so on.

A simplified model can be considered.

$$V_1 = \frac{U_{19} + U_{20} + U_{21} + U_{22}}{U_{23} + U_{24} + U_{25} + U_{26}}, \tag{6.4}$$

$$V_2 = \frac{U_{19} + U_{20} + U_{23} + U_{24}}{U_{21} + U_{22} + U_{25} + U_{26}}, \tag{6.5}$$

and

$$V_3 = \frac{U_{19} + U_{21} + U_{23} + U_{25}}{U_{20} + U_{22} + U_{24} + U_{26}}. \tag{6.6}$$

This can be used to construct a trivariate copula if the α_i's satisfy 6 constraints. For typographic simplicity, we relabel the 8 U_i's involved to have subscripts $1, 2, ..., 8$,

thus:

$$V_1 = \frac{U_1 + U_2 + U_3 + U_4}{U_5 + U_6 + U_7 + U_8}, \tag{6.7}$$

$$V_2 = \frac{U_1 + U_2 + U_5 + U_6}{U_3 + U_4 + U_7 + U_8}, \tag{6.8}$$

and

$$V_3 = \frac{U_1 + U_3 + U_5 + U_7}{U_2 + U_4 + U_6 + U_8}. \tag{6.9}$$

The required constraints are:

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1, \tag{6.10}$$

$$\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8 = 1, \tag{6.11}$$

$$\alpha_1 + \alpha_2 + \alpha_5 + \alpha_6 = 1, \tag{6.12}$$

$$\alpha_3 + \alpha_4 + \alpha_7 + \alpha_8 = 1, \tag{6.13}$$

$$\alpha_1 + \alpha_3 + \alpha_5 + \alpha_7 = 1, \tag{6.14}$$

$$\alpha_2 + \alpha_4 + \alpha_6 + \alpha_8 = 1. \tag{6.15}$$

Of these, the third and the sixth constraints are redundant, and consequently we have a four parameter family of copulas, parameterized by α, β, γ and δ, where $(\alpha, \beta, \gamma, \delta) \in [0, 1]^4$ subject to $0 \leq \alpha + \beta + \gamma \leq 1$ and $0 \leq 2\alpha + \beta + \gamma + \delta - 1 \leq 1$.

In terms of the parameters α, β, γ and δ we then have

$$\alpha_1 = \alpha, \tag{6.16}$$

$$\alpha_2 = \beta, \tag{6.17}$$

$$\alpha_3 = \gamma, \tag{6.18}$$

$$\alpha_4 = 1 - \alpha - \beta - \gamma, \tag{6.19}$$

$$\alpha_5 = \delta, \tag{6.20}$$

$$\alpha_6 = 1 - \alpha - \beta - \delta \tag{6.21}$$

$$\alpha_7 = 1 - \alpha - \gamma - \delta \tag{6.22}$$

$$\alpha_8 = 2\alpha + \beta + \gamma + \delta - 1. \tag{6.23}$$

Simplified sub-models are obtainable by setting some of α_i's equal to 0. For example, if we set $\alpha_2 = \alpha_3 = \alpha_5 = 0$ we arrive at the following one parameter model;

$$V_1 = \frac{U_1 + U_4}{U_6 + U_7 + U_8}, \tag{6.24}$$

$$V_2 = \frac{U_1 + U_6}{U_4 + U_7 + U_8}, \tag{6.25}$$

and

$$V_3 = \frac{U_1 + U_7}{U_4 + U_6 + U_8}. \tag{6.26}$$

This can then be used to construct a one parameter family of tri-variate copulas. In the model we have :

$$U_1 \sim \Gamma(\alpha, 1), \quad U_4 \sim \Gamma(1-\alpha, 1), \quad U_6 \sim \Gamma(1-\alpha, 1),$$
$$U_7 \sim \Gamma(1-\alpha, 1), \quad U_8 \sim \Gamma(2\alpha-1, 1),$$

where $\alpha \in [0.5, 1]$.

The corresponding copulas are then of the form:

$$X_1 = \frac{U_1 + U_4}{U_1 + U_4 + U_6 + U_7 + U_8}, \tag{6.27}$$

$$X_2 = \frac{U_1 + U_6}{U_1 + U_4 + U_6 + U_7 + U_8}, \tag{6.28}$$

and

$$X_3 = \frac{U_1 + U_7}{U_1 + U_4 + U_6 + U_7 + U_8}. \tag{6.29}$$

Needless to say, a family of trivariate copulas with only one parameter will be of limited utility for modeling the wide spectrum of possible dependence links among three variables. Undoubtedly, it will be advisable to use families of copulas with 2, 3 or more parameters in practice. It is unlikely that the full 20 parameter model, (6.1)-(6.3) with 26 parameters with 6 constraints, would be utilized. The four parameter family of copulas, (6.7)-(6.9) might be a reasonable first choice, though one could argue that 4 is still perhaps too low to adequately reflect the possibly complicated trivariate dependence structure of some three dimensional data sets.

For data sets of dimension four or more, the full models of gamma based copulas will have truly enormous dimensional parameter spaces (72 for four dimensions, 232 for five dimensions,etc.). Parsimonious selection of submodels, by setting many parameters equal to 0, will be essential if the models are to have utility.

7 Application

Here, we consider a Brazilian HIV data set to illustrate the feasibility of use of some of the BK copula based models developed in this article. This data set was originally used by Louzada et al. (2012). According to Louzada et al. (2012), opportunistic infections are important causes of morbidity and mortality in people with Human immunodeficiency virus (HIV), and deaths and hospitalizations are the major events resulting from these infections. In general, apart from infections acquired in living outside the hospital when hospitalized, the HIV-infected individual is found to be susceptible to hospital infections by his possible poor health condition. Some of these risk factors for nosocomial infections, particularly those of nosocomial bloodstream are the use of invasive devices, antibiotic exposure and length of stay in hospital (Tumbarello et al., 1998). In a nutshell, the general health condition of the patient admitted to the service as well as the occurrence of an infection acquired during hospitalization, can further extend his length of stay in hospital. Furthermore, it is quite reasonable to consider the fact that the HIV-infected patient may require one or more hospital admissions and length of stay in this environment can contribute significantly to serious complications, which may lead to death of the patient. Also, the timing of a given hospitalization may somehow depend on the time the patient been hospitalized previously.

Consequently, there is a need to study the length of stay at the hospital corresponding to two different times. For a detailed description of the data, an interested reader is referred to Louzada et al. (2012). This data set contains information on patients (older than 18 years of age) two different stages of admission with HIV seen at the Servicio de Doencas Infecciosas e Parasitarias (DIP), Universidade Federal do Triangulo Mineiro (UFTM), Brazil, diagnosed with HIV between January 1996 and December 1998. Here, we consider one application of six of the bivariate Kumaraswamy copula models to this data. We consider two variables:

- X_1 : The proportion of timely admitted patients (first admission, preferably at the first stage of HIV, i.e., chance of partial/and or complete cure to HIV) for each month during the time interval January 1996-December 1998.

- X_2 : The proportion of patients (for the second stage/second admission) who successfully came out of the first round of treatment and subsequently enrolled to this phase during the same time period as in X_1.

  So we have 36 such paired values of (X_1, X_2).

The data is fitted to those six selected BK (copula based construction) models. In the table below, we provide the p-values for several goodness-of-fit tests based on certain statistics described in Genest et al. (2006).

Table 1. Goodness of fit statistics for copula based BK models for the Brazilian HIV data.

Copula on which the BK model is based	p-value
Nested AN (Equations (4.7)-(4.8)) (BK model has 5 parameters)	0.7310
Nested AN (Equations (4.5)-(4.6)) (BK model has 6 parameters)	0.7568
Marshall - Olkin (Equation (5.4)) (BK model has 5 parameters)	0.1689
Gumbel-Houggard (Equation (5.3)) (BK model has 5 parameters)	0.0065
Ali-Mikhail-Haq (Equation (5.2)) (BK model has 5 parameters)	0.0002
FGM (Equation (5.1)) (BK model has 5 parameters)	0.3205

The p-values reported in Table 1 are based on $N = 10000$ repetitions of the parametric bootstrap procedure discussed in Genest et al. (2006). From Table 1, it appears that nested A-N copula based bivariate Kumaraswamy with 6 and 5 parameters provide better fits when compared with other BK copula models, and that, as expected, the 6 parameter model is best of all.

8 Comments

The example investigated in the last section indicates that AN-copula based Kumaraswamy models do have some potential for fitting bounded data sets. A more detailed and extensive comparative study is clearly called for. Plans for such an investigation are underway. It is, of course, not to be expected that one copula-based bivariate Kumaraswamy model will be uniformly best. Which model is best will usually depend on the nature of the particular data set at hand. Hopefully, it will be possible to give some guidance to users of such models. The survey of available non-AN copulas which might be used made in this article is far from complete and, in some instances, bivariate Kumaraswamy models not discussed here will prove to be preferable.

References

[1] Arnold, B.C. and Ghosh, I. (2016 a). Some alternative bivariate Kumaraswamy models. *To appear in Communications in Statistics:Theory and Methods, DOI:10.1080/03610926.2016.1208244*.

[2] Arnold, B.C. and Ghosh, I. (2016 b). Bivariate Beta and Kumaraswamy models developed using the Arnold-Ng bivariate beta distribution. *Submitted*.

[3] Arnold, B. C., and Ng, H. K. T (2011). Flexible bivariate beta distributions. *Journal of Multivariate Analysis*, 102, 1194-1202.

[4] Balakrishnan, N., and Lai, C.D. (2009). Continuous Bivariate Distributions, Second edition. Springer, New York.

[5] Barreto-Souza, W. and Lemonte, A.J. (2013). Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes. *Statistics*, 47, 1-22.

[6] Chen, X., and Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Can. J. Statist.,* 33, 389-414.

[7] Cook, R. D., and Johnson, M.E. (1981). A family of distributions for modeling non-elliptically symmetric multivariate data. *J. Roy. Statist. Soc. Ser. B,* 43, 210-218.

[8] Frees, E. W., and Valdez, E. A. (1998). Understanding relationships using copulas. *North Am. Act. J.* 2, 1-25.

[9] Genest, C., Quessy, J.R. and B. Remillard, B. (2006). Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation. *Scandinavian Journal of Statistics,* doi:10.1111/j.1467-9469.2006.00470.

[10] Ghosh, I., and Ray, S. (2016). Some alternative bivariate Kumaraswamy type distributions via copula with application in risk management. *Journal of Statistical Theory and Practice, DOI: 10.1080/15598608.2016.1215943.*

[11] Ghosh, I. (2016). Some convolutions on modified FGM copula and it's role in constructing bivariate Kumaraswamy models. *Submitted.*

[12] Klugman, S. A., and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insur. Math. Econ.* 24, 139-148.

[13] Kumaraswamy, P. (1980). Generalized probability density-function for double-bounded random-processes. *Journal of Hydrology*, 462, 79–88.

[14] Louzada,F., Adriano K. Suzuki, A.K., Cancho, V.G., Prince, F.L., and Pereira, G.A. (2012).The Long-Term Bivariate Survival FGM Copula Model: An Application to a Brazilian HIV Data. *Journal of Data Science,* 10, 511–535.

[15] Nelsen, R.B. (2006). *An Introduction to Copulas.* Springer-Verlag, New York.