Local Dependence in Bivariate Copulae with Beta Marginals

Dependencia local en copulas bivariadas con marginales Beta

Eirini Koutoumanou[a], Angie Wade[b], Mario Cortina-Borja[c]

Population Policy and Practice Programme, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

Abstract

The local dependence function (LDF) describes changes in the correlation structure of continuous bivariate random variables along their range. Bivariate density functions with Beta marginals can be used to model jointly a wide variety of data with bounded outcomes in the (0,1) range, e.g. proportions. In this paper we obtain expressions for the LDF of bivariate densities constructed using three different copula models (Frank, Gumbel and Joe) with Beta marginal distributions, present examples for each, and discuss an application of these models to analyse data collected in a study of marks obtained on a statistics exam by postgraduate students.

Key words: Association, Beta distribution, Bivariate distribution, Copula, Correlation.

Resumen

La función de dependencia local (FDL) describe cambios en la estructura de la correlación entre dos variables aleatorias continuas sobre su recorrido conjunto. Funciones bivariadas de densidad de probabilidad con densidades marginales Beta pueden utilizarse apar modelar conjuntamente una amplia variedad de variables respuesta acotadas en el intervalo (0, 1), por ejemplo proporciones.

En este artículo obtenemos expresiones para la FDL de densidades bivariadas utilizando tres modelos de cópulas (Frank, Gumbel y Joe) con densidades marginales Beta, presentamos ejemplos para cada una de estas funciones, y discutimos una aplicación de estos modelos al análisis de datos recolectados en un estudio de calificaciones para un examen de estadística aplicado a estudiantes de postgrado.

Palabras clave: asociación, distribución Beta, cópula, correlación, función de distribución bivariada.

[a]Miss. E-mail: e.koutoumanou@ucl.ac.uk

[b]PhD. E-mail: a.wade@ucl.ac.uk

[c]PhD. E-mail: m.cortina@ucl.ac.uk

1. Introduction

Interest often lies in studying bivariate bounded distributions, for example where both random variables (rv's) are rates or proportions limited from 0 to 1 and these can be efficiently modelled using bivariate Beta distributions. There are many fields of application for such joint models, e.g. mathematics and language exam marks of students, results of psychological tests applied to matched groups of patients receiving intervention and placebo, the percentiles of height and weight, or the proportions of household income spent on food and on heating.

There are several models for bivariate Beta distributions: some are derived from transformations of three standard (Olkin & Liu 2003), non-central and five (Gupta, Orozco-Castañeda & Nagar 2011) Gamma-distributed rv's; others arise from the relations between the Beta, F and skew-t distributions (Jones 2002), (El-Bassiouny & Jones 2009).

Sometimes there is no desire to impose constraints on the data-generating mechanism, such as those driven by transformations of Gamma densities, and alternative processes are required. Thus, another possibility is provided by the class of bivariate distributions with Beta marginals constructed via copula functions. Copulae are functions that join univariate cumulative distribution functions (cdf) into a multivariate cdf (Sklar 1959). Let $F_1(x_1)$ and $F_2(x_2)$ denote the univariate marginals of beta-distributed rv's X_1 and X_2, respectively. Sklar's theorem establishes that if H is a two-dimensional cdf with marginals F_1 and F_2, then there exists uniquely a copula $C : [0,1] \times [0,1] \to [0,1]$, such that for all rv's X_1 and X_2 in $[-\infty, \infty]$:

$$H\ (x_1, x_2; \theta) = C\ (F_1(x_1), F_2(x_2))\,.$$

The copula parameter θ defines the degree of association between the marginals and it is well-known that it is directly associated with Kendall's correlation coefficient (Schweizer & Wolff 1981), as follows:

$$\tau = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H\ (x_1, x_2; \theta)\ \partial H\ (x_1, x_2; \theta) - 1.$$

Moreover, a scalar measure of correlation, e.g. Pearson's r, is often not enough to adequately describe the dependence structure of bivariate rv's (X_1, X_2) with continuous joint density function f. An interesting review paper on copula and dependence structures is (Escarela & Hernández 2009).

Local dependence measures allow a thorough exploration of the nature of the joint variation by analysing changes in the strength of the association along the range of (X_1, X_2). There are several models and graphical representations of the notion of local dependence, for example local dependence maps (Jones & Koch 2003), local Gaussian correlation (Tjøstheim & Hufthammer 2013), Kendall plots (Genest & Boies 2003) and chi-plots (Fisher & Switzer 1985). Here we focus on the definition of the local dependence function (LDF) as given by Holland & Wang (1987):

$$\gamma\left(x_1, x_2\right) = \frac{\partial^2 \ln f\left(x_1, x_2\right)}{\partial x_1 \; \partial x_2}. \tag{1}$$

The $\gamma\left(x_1, x_2\right)$ LDF calculates the rate of change of the natural logarithm of the bivariate density at every combination of the x_1 and x_2 values. The LDF can be seen as a localization of the Pearson correlation coefficient r for $\left(X_1, X_2\right)$ (Jones 1996), and measures the strength of the association between X_1 and X_2 in a neighbourhood of any point $\left(x_1, x_2\right)$ in the domain of f. For independent rv's γ is uniformly 0, and it is constant only for the few distributions identified by (Jones 1998), among them the bivariate normal with value $\gamma = r / \left(1 - r^2\right)$.

The class of bivariate models with Beta marginals provides an interesting framework in which to explore the role of the LDF in revealing bivariate structures because the Beta distribution is extremely flexible and can produce probability density functions (pdf's) in many shapes, e.g. U- and J-shaped, symmetric, and even uniform. Some work on this has been done by (Gupta, Kirmani & Srivastava 2010); see the LDF formula in equations 11 and 20 for the Farlie–Gumbel–Morgenstern (FGM) (Morgernstern 1956, Gumbel 1960, Fairlie 1960) and Ali–Mikhail–Haq (AMH) (Ali, Mikhail & Haq 1978 and Kumar 2010) copulae with Beta marginals, respectively. We aim to extend these LDF formulae to more bivariate copulas with Beta marginals that have not been previously studied. The LDF, as given in equation (1), is directly related to the density of the bivariate distribution itself via its moments as shown by Jones (1996), hence it is easy to construct when the bivariate distribution is fully specified. We hope to improve research practice in exploring bivariate associations by providing the LDF for three bivariate distributions obtained via coupling Beta marginals and focusing on the interpretation of their graphical displays.

The copula models featured in this paper were chosen based on their widespread usage in the multivariate research field, the ease of construction of their LDF as well as their ability to capture a wide range of dependency structures. We focus on three models from the Archimedean family (Nelsen 2013): Frank (1979), Gumbel (1960) and Joe (1993, 1997).

In the next section, we provide expressions of the LDF for the three bivariate copulas with Beta marginals (as obtained via the computational software Mathematica version 10 (Wolfram Research Inc 2014) using the MathStatica extension (Rose & Smith 2002)) along with illustrations of the copula density and the LDF for each with varying parameters (using R Development Core Team (2007), version 3.1.2) to facilitate the reader's understanding and interpretation of the LDF figures. In section 3, and for the purposes of completeness, we discuss the graphical interpretation of the results for the FGM and AMH copulas with Beta marginals as presented by Gupta et al. (2010).

There are very few published applications of the LDF to real data despite its great potential for its use on non-Normal bivariate associations that researchers often deal with. Usually analysis of such associations is restricted simply to a nonparametric scalar correlation coefficient, such as Spearman's or Kendall's. In section 4, we fit and compare copula models to obtain the maximum likelihood

estimates (Yan 2007) of the parameters of the bivariate pdf with Beta marginals to a dataset from the field of statistical education and we demonstrate the usefulness of the derived LDF formulations as an alternative to scalar correlation coefficients.

2. Copula-Defined Bivariate Distributions with Beta Marginals

Let X_1, X_2 be univariate random variables each with a univariate Beta distribution with shape parameters $a_i, b_i \geq 0$, $i = 1, 2$ respectively, i.e.:

$$f_i\left(x_i\right) = \frac{x_i^{a_i-1}\left(1-x_i\right)^{b_i-1}}{B\left(a_i, b_i\right)} \text{ and } F_i\left(x_i\right) = \frac{B\left(x_i, a_i, b_i\right)}{B\left(a_i, b_i\right)}$$

where $B\left(x; a, b\right) = \int_0^x t^{a-1}\left(1-t\right)^{b-1}\partial t$ denotes the incomplete beta function, and $B\left(a, b\right) = B\left(1, a, b\right)$. We denote the survival function as $\bar{F}\left(x\right) = 1 - F\left(x\right)$.

In each copula section below each LDF is followed by two pairs of graphs displaying the contours of the copula density function on the left and the LDF on the right for different marginal and dependence parameters. The parameters of the bivariate distribution (a_i, b_i, and θ) differ between each copula for the first pair of graphs (Figures 1, 3 and 5), but are the same for the second set of graphs (Figures 2, 4 and 6). This illustrates how the same copula function can lead to markedly different bivariate structures for varying marginal and dependence parameters as well as how the same parameters can lead to markedly different bivariate associations for differing copula functions. The contours are slices of a distribution drawn at certain density levels; in these examples the following values are shown: 0, 0.2, 1, 2, 3, 4, 5, 10, 15 and 100. These represent density levels on the left hand side graph and local dependence levels on the right hand side graphs. For both graphs, the x and y axes represent the two rv's, X_1 and X_2 respectively.

2.1. Frank Copula

A Frank copula is constructed via the following combination of univariate marginals $F_1\left(x_1\right)$ and $F_2\left(x_2\right)$:

$$F\left(x_1, x_2\right) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta F_1\left(x_1\right)} - 1\right)\left(e^{-\theta F_2\left(x_2\right)} - 1\right)}{e^{-\theta} - 1}\right)$$

where $\theta \in (-\infty, \infty), \theta \neq 0$ to ensure absolute continuity; if $\theta = 0$ then X_1, X_2 are independent (as before).

The density of the copula with Beta marginals can be written as follows and this corresponds to a five-parameter family of bivariate distributions with Beta marginals (for $i = 1, 2$):

$$f\left(x_1, x_2\right) = \theta\left(e^\theta - 1\right)\left[\prod_i f_i\left(x_i; a_i, b_i\right)\right] \frac{e^{\theta\left(1 + \sum F_i\left(x_i\right)\right)}}{\left(\sum_i e^{\theta\left(1 + F_i\left(x_i\right)\right)} - e^\theta \sum F_i\left(x_i\right) - e^\theta\right)^2}$$

The LDF of this copula is equal to:

$$\begin{aligned}
\gamma\left(x_1, x_2\right) &= 2\,\theta^2\left(e^\theta - 1\right)\left[\prod_i f_i\left(x_i; a_i, b_i\right)\right] \\
&\quad \times \frac{e^{\theta\left(1 + \sum_i F_i\left(x_i\right)\right)}}{\left(\sum_i e^{\theta\left(1 + F_i\left(x_i\right)\right)} - e^\theta \sum F_i\left(x_i\right) - e^\theta\right)^2} \\
&= 2\,\theta\, f\left(x_1, x_2\right)
\end{aligned}$$

Notice that the latter has the same sign as θ all over the unit square.

Figure 1 presents the pdf and LDF contours of a Frank copula with given parameters.

FIGURE 1: pdf and LDF of Frank copula with marginal and dependence parameters: $a_1 = 5, b_1 = 2; a_2 = 5, b_2 = 2, \theta = -12$.

Both the pdf and the LDF follow the same general pattern with the only difference being the fact that the local dependence function expands more widely to accommodate for pairs of x_1 and x_2 that have low bivariate density but are locally associated with respect to the whole range of values. Notice the negative values of the LDF contours, as a result of the negative dependence parameter.

The following set of graphs shows a Frank copula with changed marginal and dependence parameters. Notice that the general shape of the LDF is very similar to that of the pdf, as in the previous example, i.e. local dependence values mirror very well the density values around the same areas of the bivariate relationship.
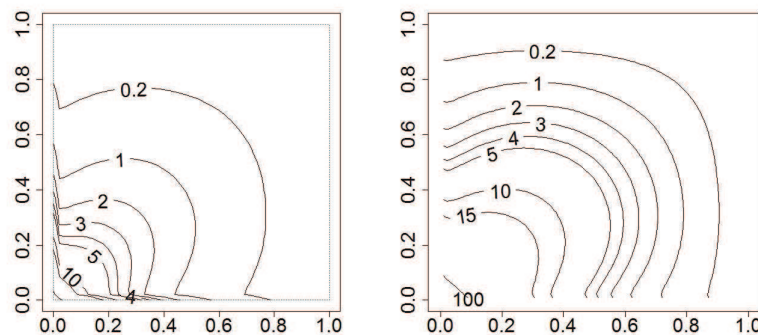
FIGURE 2: pdf and LDF of Frank copula with marginal and dependence parameters: $a_1 = 0.8, b_1 = 3, a_2 = 0.8, b_2 = 3, \theta = 3$.

2.2. Gumbel Copula

The Gumbel copula is defined as follows, where $\theta \in [1, \infty)$:

$$F(x_1, x_2) = \exp\left(-\left[(-\ln F_1(x_1))^\theta + (-\ln F_2(x_2))^\theta\right]^{1/\theta}\right)$$

The density of the copula when the two marginals are Beta distributions and its LDF can be written as (for $i = 1, 2$):

$$
\begin{aligned}
f(x_1, x_2) &= \prod_i f_i(x_i; a_i, b_i) \prod_i F_i^{-1}(x_i; a_i, b_i) \prod_i (L_{x_i})^{\theta-1} \\
&\times \left(\theta - 1 + \sum_i (L_{x_i})^\theta\right)^{\frac{1}{\theta}} \left(\sum_i (L_{x_i})^\theta\right)^{\frac{1}{\theta}-2} \exp\left(-\sum_i (L_{x_i})^\theta\right)^{\frac{1}{\theta}}
\end{aligned}
$$

where $L_{x_i} = -\ln F_i(x_i)$, and

$$
\begin{aligned}
\gamma(x_1, x_2) &= (\theta - 1)\prod_i f_i(x_i; a_i, b_i) \prod_i F_i^{-1}(x_i; a_i, b_i) \\
&\times \left(\sum_i (L_{x_i})^\theta\right)^{\frac{1}{\theta}-2} \prod_i (L_{x_i})^{\theta-1} \\
&\times \left\{\theta(2\theta - 1)\left[(\theta-1)(2 + 5\theta(\theta-1)) + \left(\sum_i (L_{x_i})^\theta\right)^2\right]\right. \\
&+ \left.\left(\sum (L_{x_i})\theta\right)^3\right\}
\end{aligned}
$$

In contrast with the previous copula, there is no linear relationship between the dependence parameter θ and the LDF for this copula.

The pdf graph of the copula shown in Figure 3 may give the impression of a similar positive association for most of their joint range. However, the LDF graph provides a more thorough view of the dependence. The correlation is maximal across the main positive diagonal whilst it decreases rather quickly off diagonal and becomes minimal along the (0, 1) and (1, 0) axes.
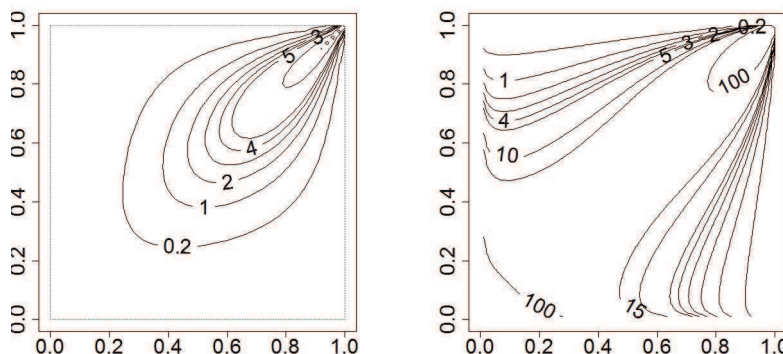

FIGURE 3: pdf and LDF of Gumbel copula with marginal and dependent parameters: $a_1 = 5, b_1 = 2, a_2 = 5, b_2 = 2, \theta = 2$.

Figure 4 shows another example of the density and dependence functions of a Gumbel copula. The pdf emphasises the high density at the bottom left corner of the bivariate distribution, whereas the LDF shows a strong correlation along the central part of the distribution.
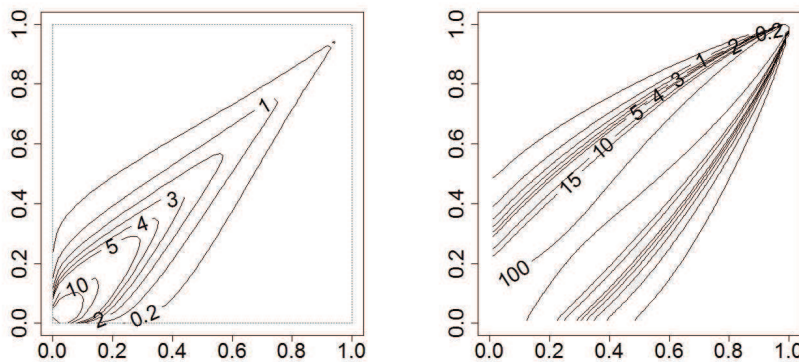

FIGURE 4: pdf and LDF of Gumbel copula with marginal and dependence parameters: $a_1 = 0.8, b_1 = 3, a_2 = 0.8, b_2 = 3, \theta = 3$.

2.3. Joe Copula

The Joe copula with Beta marginals yields the following bivariate distribution, where $\theta \in (1, \infty)$:

$$F(x_1, x_2) = 1 - \left[(1 - F_1(x_1))^\theta + (1 - F_2(x_2))^\theta - (1 - F_1(x_1))^\theta (1 - F_2(x_2))^\theta \right]^{\frac{1}{\theta}}$$

The density of the copula when the two marginals are Beta distributions and its LDF can be written, respectively, as follows:

$$
\begin{aligned}
f(x_1, x_2) \quad &= \quad f(x_1)\, f(x_2)\, \left(\bar{F}(x_1)\right)^{\theta-1} \left(\bar{F}(x_2)\right)^{\theta-1} \\
&\times \quad \left[\left(\bar{F}(x_1)\right)^\theta - \left(\bar{F}(x_2)\right)^\theta \left(\bar{\bar{F}}_{x_1}\right) \right]^{\frac{1}{\theta}-2} \left[\theta - \bar{\bar{F}}_{x_1}\, \bar{\bar{F}}_{x_2} \right],
\end{aligned}
$$

where $\bar{\bar{F}}_{(x_i)} = \left(\bar{F}(x_i)\right)^\theta - 1$ for $i = 1, 2$ and:

$$
\begin{aligned}
\gamma(x_1, x_2) \quad &= \quad f(x_1)\, f(x_2)\, \theta\,(\theta-1)\, [F(x_1)\, F(x_2)]^{\theta-1} \\
&\times \frac{-2\,\theta^2 - \left(\bar{\bar{F}}_{x_1}\right)^2 \left(\bar{\bar{F}}_{x_2}\right)^2 + \theta\, \bar{\bar{F}}_{x_1}\, \bar{\bar{F}}_{x_2}\, \left[3 - \left(\bar{F}(x_1)\right)^\theta + \bar{\bar{F}}_{x_1}\, \bar{\bar{F}}_{x_2} \right]}{\left\{ \left[\left(\bar{F}(x_1)\right)^\theta - \left(\bar{F}(x_1)\right)^\theta \bar{\bar{F}}_{x_1} \right] \left[\theta - \bar{\bar{F}}_{x_1} \bar{\bar{F}}_{x_2} \right] \right\}^2}
\end{aligned}
$$

As with the Gumbel copula, the overall sign of the LDF cannot be determined directly from the sign of θ.

In Figure 5, density is highest at the top right corner whilst the remaining corners have lower peaks. The highest levels of local dependence are found in similar locations on the LDF graph too, i.e. all corners, with the top right corner having the highest level of local dependence, i.e. large values of both X_1 and X_2 rv's are most highly correlated.
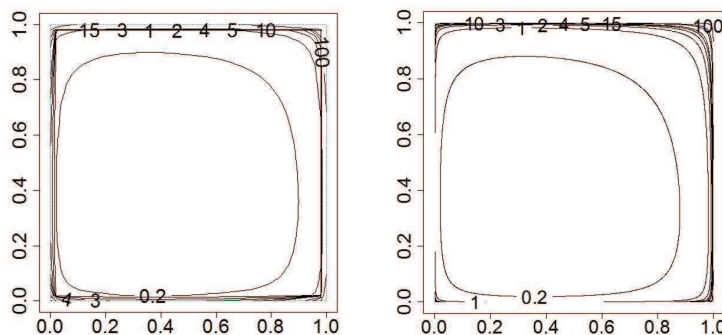

FIGURE 5: pdf and LDF of Joe copula with marginal and dependence parameters: $a_1 = 0.5, b_1 = 0.1, a_2 = 0.5, b_2 = 0.1, \theta = 1.5$.

The bivariate copula shown in Figure 6 has the same parameters as the second example of the previous two copulas, Frank (Figure 2) and Gumbel (Figure 4). This illustrates the flexibility of various copula models to capture changing

bivariate associations/structures given the same parameters. This is mirrored in the LDF graphs too, capturing local dependence at all sets of neighbouring points. The LDF in Figure 6 is an example of how the same amount of local dependence is found at two different areas of the bivariate plot, i.e. for low values of X_1 and X_2 and for values of X_1 and X_2 within the range of $(0.4, 1)$.
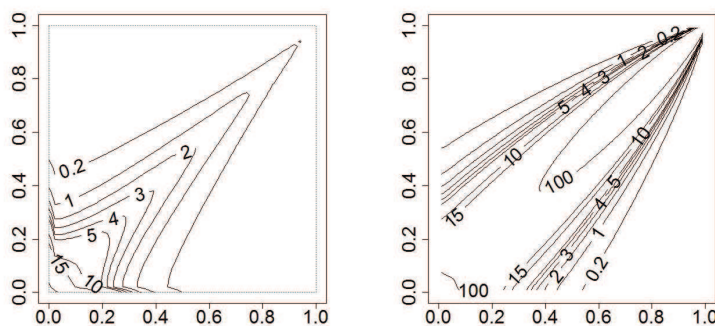

FIGURE 6: pdf and LDF of Joe copula with marginal and dependence parameters: $a_1 = 0.8, b_1 = 3, a_2 = 0.8, b_2 = 3, \theta = 3$.

3. More Examples and Results

For completeness, we present here examples of the FGM and AMH copulas based on the expressions presented by Gupta et al. (2010) in Figures 7 and 8 respectively. Notice that, although the general shapes of the two densities are similar, the LDF's are markedly different. The LDF of this FGM model has the highest correlation for x_1 values higher than 0.4 and for low values of x_2 and similarly for x_2 values higher than 0.4 for low values of x_1, whilst this pattern was not evident on the density graph. In contrast, the AMH copula reaches its highest local dependence close to the bottom left corner of the graph, where the highest density is also observed.
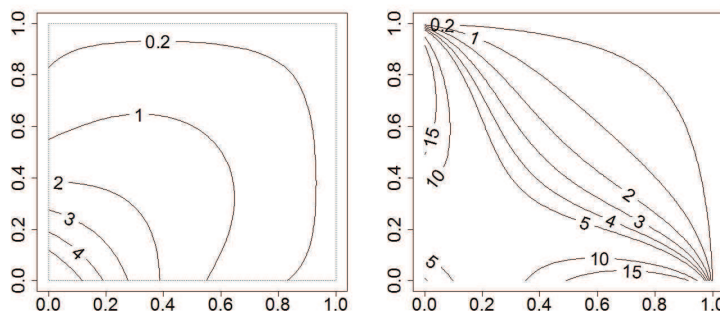

FIGURE 7: pdf and LDF of FGM copula with marginal and dependence parameters: $a_1 = 1, b_1 = 2, a_2 = 1, b_2 = 2, \theta = 0.75$.
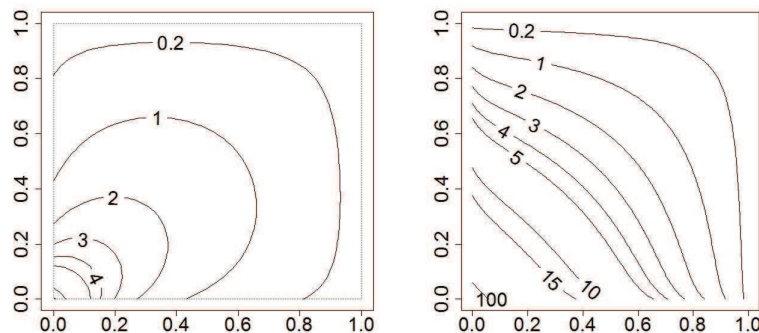
FIGURE 8: pdf and LDF of AMH copula with marginal and dependence parameters: $a_1 = 1, b_1 = 2, a_2 = 1, b_2 = 2, \theta = 0.75$.

An interesting result for the FGM copula expresses Pearson's r for this bivariate model as a function of its five parameters. Since this bivariate distribution is constructed as an FGM copula its joint moments reduce to the product of univariate integrals in each variable as shown by D'Este (1981) and Gupta & Wong (1985), hence

$$r = \theta \prod_{i=1}^{2} \left[\frac{\frac{2\,B\,(2\,a_i+1,b_i)\,_3F_2\,(\{a_i,\ 2\,a_i+1,\ 1-b_i\};\{a_i+1,\ 2\,a_i+b_i+1\};1)}{B\,(a_i,b_i)\,B\,(a_i+1,b_i)} - 1}{\sqrt{\frac{a_i\,b_i}{a_i+b_i+1}}} \right]$$

where $_3F_2$ is the generalized regularized hypergeometric function (El-Bassiouny & Jones 2009). It was shown by Schucany, Parr & Boyer (1978) that the absolute value of the correlation coefficient for any FGM copula is less than or equal to $1/3$ together with examples of particular marginals with correlations smaller than $1/3$. It is easy to see that this bound is reached for this bivariate distribution, e.g. if $a_1 = b_1 = a_2 = b_2 = 1$ then $r = \theta/3$. If all the four parameters of the univariate Beta distributions in the copula are equal, r takes values between $\theta/4$ (all 0) and $\theta/3$ (all 1) and then decreases very little to stay just above $0.318\,\theta$ for parameter values much larger than 1. We were only able to reach this result for the FGM copula due to the ease of the calculation of the moments of its bivariate expression; this was not possible for the other bivariate copulas presented in this paper.

4. Application

Having calculated the cdfs and LDFs of three copula models with Beta marginals, we analyse the association between postgraduate students' grades for a practical SPSS task and a set of multiple-choice questions (MCQ) together comprising a statistics module exam. The data were collected as part of an MSc Statistics module ran at University College London by the first two authors in 2013 ($n = 116$

students), and are shown in Figure 9. Both variables are expressed as proportions between 0 and 1 and the median grade for the SPSS task is 0.72 and for the MCQ questions is 0.51, with $\hat{r} = 0.67$ and $\hat{\tau} = 0.54$. Additionally, the estimates of multivariate skewness (m_1) and kurtosis (m_2) defined by Mardia (1970) have values of 1.5 and 10.7, respectively, providing evidence of bivariate non-Normal distribution when compared with the threshold values of 0 and $d / (d + 2) = 0.5$, where d is the dimensionality of the dataset, which indicate multivariate Normality (Mardia 1974).
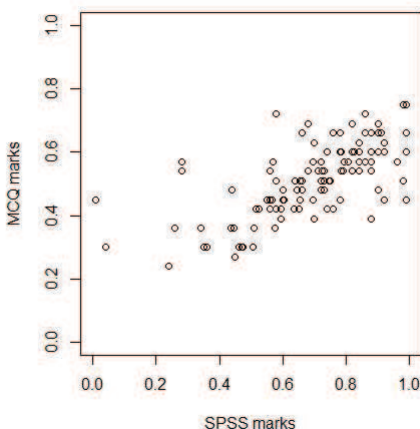

FIGURE 9: Scatterplot of students' marks for SPSS and MCQs.

By fitting a bivariate Normal distribution, we obtain bivariate mean $= \begin{bmatrix} 0.70 \\ 0.51 \end{bmatrix}$ and bivariate covariance matrix $= \begin{bmatrix} 0.040 & 0.015 \\ 0.015 & 0.013 \end{bmatrix}$. However, this distribution is neither a natural choice for this data set due to the bounded nature of the variables nor does it provide a good fit (BIC $= -266$) compared to other models that were fitted, as shown in Table 1.

For this application, we fitted the Frank, Gumbel and Joe copulas with the use of univariate Beta marginals for the two related proportions. The use of the AMH and FGM copulas is not possible for this dataset, as the estimated τ is outside the allowed bounds of $(-0.18, 0.33)$ and $(-0.22, 0.22)$ respectively for each of these copula models.

TABLE 1: AIC and BIC of fitted models

Copula	AIC	BIC
Frank	-302.02	-288.26
Gumbel	-281.24	-267.47
Joe	-263.56	-249.79

According to the BIC and AIC values (Kuha 2004), the Frank copula provides the best fit to this dataset. The left hand side graph of Figure 10 presents the

contours of the bivariate copulae based on the maximum likelihood estimates of the parameters of the each of the Beta marginals and the dependence parameter θ (all were found significant at 5%). The right hand side graphs show the contours of the associated LDF.
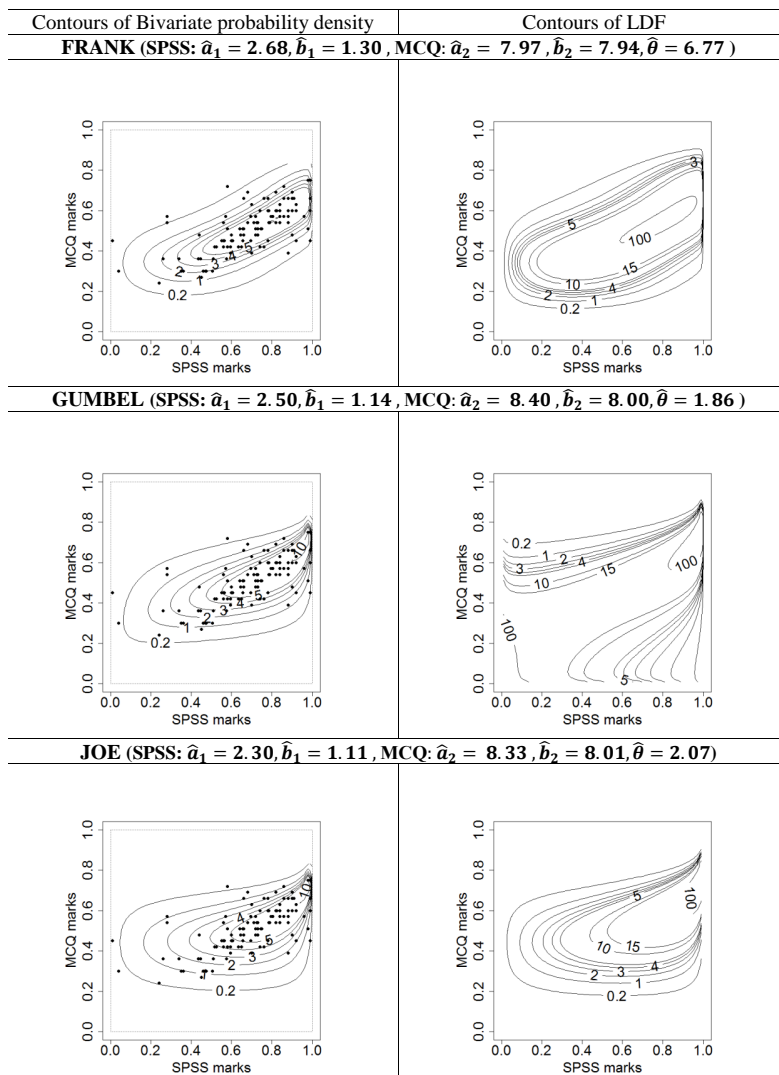

FIGURE 10: Frank, Gumbel and Joe copula bivariate distribution for data on students' marks.

The fitted pdf evaluates how densely observations occur within a certain area but does not tell us anything about the association (not necessarily linear) of those points within the same area, which is what the LDF does. The pdf of the Frank copula model shows higher density around marks of (0.6,0.9) and (0.4,0.6) as well

as a wider spread of the values for the lower marks, hence wider contours. This is additionally emphasised in the LDF where the highest local dependence value is located at the exact same range of values with lower local dependence values found for the remaining ranges. The LDF provides a more comprehensive description of the relationship between the two types of grades, which would have otherwise been summarised merely via a constant Kendall's correlation coefficient of $\tau = 0.54$.

Focusing more on the LDF, what we know from years of teaching experience is apparent in this sample; practical tasks are able to boost up student's overall marks, where students can mark relatively well on the practical task (achieve marks between 0.6 and 0.9) whilst they only achieve a pass on average on the theoretical questions (marks between 0.4 and 0.6).

The Gumbel and Joe copulae fail to capture the evident dependency between the SPSS and MCQ marks. The Gumbel copula puts more emphasis on the peripheral points of this graph, the two extreme areas of the bivariate plot, and entirely misses the strong dependency at the most dense areas of the scatterplot. The Joe copula adopts a flatter association between the two variables; with marks for the software-based question between 0.2 and 0.9 being associated with marks ranging from 0.3 to 0.6 for the non-practical questions, suggesting that students can perform rather well at the SPSS task whilst their theoretical marks might not be as high.

5. Discussion

The flexibility of individual copula types to cover a wide range of bivariate distributions with Beta marginals is emphasised via the use of different parameter sets (contrast Figures 1 and 2, 3 and 4 and 5 and 6). To show the difference between copula models note that we have used the same parameter sets for Figures 2, 4 and 6. It is important for researchers to realise that correlation coefficients do not always convey the relationship between numerical variables in the best possible way. Summarising the entire correlation structure in a single constant value does not account for changing the strength of the association across the joint range of x_1 and x_2. The local dependence function overcomes this problem by producing a detailed graphical display of the association between x_1 and x_2 across all values. The LDF can produce distinctly different graphs for bivariate density functions that look very much alike as demonstrated in the contrasting Figures 7 and 8.

We have presented three LDF formulae, which can be easily programmed to produce LDF graphs for bivariate scenarios where Beta marginals are a relevant and good choice for the univariate density functions.

Our results could be extended to incorporate covariates, hence LDF could be then drawn conditional on selected predictor values. Additionally, the same principles used here could be applied to marginals other than univariate Beta such as the multivariate skew-normal distribution defined by Azzalini & Dalla Valle (1996). Such extensions would lead to even greater scope for the application of

LDF in many fields where bivariate rv's are analysed thus enhancing researchers' understanding of local dependence structures.

6. Conclusions

We have used a novel approach to explore bivariate associations between rv's via the combination of the notion of local dependence and copulas. We anticipate that this work will facilitate better research within the field of multivariate dependence structures.

Acknowledgments

References

Ali, M. M., Mikhail, N. & Haq, M. S. (1978), 'A class of bivariate distributions including the bivariate logistic', *Journal of Multivariate Analysis* **8**(3), 405–412.

Azzalini, A. & Dalla Valle, A. (1996), 'The multivariate skew-normal distribution', *Biometrika* **83**(4), 715–726.

D'Este, G. M. (1981), 'A Morgenstern-type bivariate gamma distribution', *Biometrika* **68**(1), 339–340.

El-Bassiouny, A. & Jones, M. (2009), 'A bivariate F distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate beta and t distributions', *Statistical Methods and Applications* **18**(4), 465–481.

Escarela, G. & Hernández, A. (2009), 'Modelling random couples using copulas', *Revista Colombiana de Estadística* **32**(1), 33–58.

Fairlie, D. (1960), 'The performance of some correlation coefficients for a general bivariate distribution', *Biometrika* **18**(3/4), 307–323.

Fisher, N. & Switzer, P. (1985), 'Chi-plots for assessing dependence', *Biometrika* **72**(2), 253–265.

Frank, M. (1979), 'On the simultaneous associativity of $F(x, y)$ and $x+y−F(x, y)$', *Aequationes Matematicae* **19**(1), 194–226.

Genest, C. & Boies, J. (2003), 'Detecting dependence with Kendall plots', *The American Statistician* **57**(4), 275–284.

Gumbel, E. (1960), 'Distributions des valeurs extrêmes en plusieurs dimensions', *Publications de l'Institut de statistique de l'Université de Paris* **9**, 171–173.

Gupta, A., Orozco-Castañeda, J. & Nagar, D. (2011), 'Non-central bivariate Beta distribution', *Statistical Papers* **52**(1), 139–152.

Gupta, A. & Wong, C. (1985), 'On three and five parameter bivariate Beta distributions', *Metrika* **32**, 85–91.

Gupta, R., Kirmani, S. & Srivastava, H. (2010), 'Local dependence functions for some families of bivariate distributions and total positivity', *Applied Mathematics and Computation* **216**(4), 1267–1279.

Holland, P. W. & Wang, Y. J. (1987), 'Dependence function for continuous bivariate densities', *Communications in Statistics-Theory and Methods* **16**(3), 863–876.

Joe, H. (1993), 'Parametric families of multivariate distributions with given margins', *Journal of Multivariate Analysis* **46**(2), 262–282.

Joe, H. (1997), *Multivariate Models and Multivariate Dependence Concepts*, Chapman and Hall / CRC, London.

Jones, M. (1996), 'The local dependence function', *Biometrika* **83**(4), 889–904.

Jones, M. (1998), 'Constant local dependence', *Journal of Multivariate Analysis* **64**(2), 148–155.

Jones, M. (2002), 'Multivariate t and Beta distributions associated with the multivariate F distribution', *Metrika* **54**(3), 215–231.

Jones, M. & Koch, I. (2003), ' Dependence maps: Local dependence in practice', *Statistics and Computing* **13**(3), 241–255.

Kuha, J. (2004), 'AIC and BIC: Comparisons of assumptions and performance', *Sociological Methods & Research* **33**(2), 188–229.

Kumar, P. (2010), 'Probability distributions and the estimation of Ali-Mikhail-Haq Copula', *Applied Mathematical Statistics* **4**(14), 657–666.

Mardia, K. (1970), 'Measures of multivariate skewness and kurtosis with applications', *Biometrika* **57**(3), 519–530.

Mardia, K. (1974), 'Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies', *Sankhyā: The Indian Journal of Statistics, Series B* **36**(2), 115–128.

Morgernstern, D. (1956), 'Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies', *Mitteilungsblatt für Mathematische Statistik* **8**, 234–235.

Nelsen, R. (2013), *An Introduction to Copulas*, Springer, New York.

Olkin, I. & Liu, R. (2003), 'A bivariate beta distribution', *Statistics & Probability Letters* **62**(4), 407–412.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Rose, C. & Smith, M. (2002), *Mathematical Statistics with Mathematica*, Springer, New York.

Schucany, W., Parr, W. & Boyer, J. (1978), 'Correlation structure in Farlie-Gumbel-Morgenstern distributions', *Biometrika* **65**(3), 650–653.

Schweizer, B. & Wolff, E. (1981), 'On nonparametric measures of dependence for random variables', *The Annals of Statistics* **8**(2), 878–885.

Sklar, A. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de statistique de l'Université de Paris* **8**, 229–231.

Tjøstheim, D. & Hufthammer, K. O. (2013), 'Local Gaussian correlation: A new measure of dependence', *Journal of Econometrics* **172**(1), 33–48.

Wolfram Research Inc (2014), *Mathematica version 10*, Wolfram Research Inc, Champaign, Illinois.
*http://www.wolfram.com

Yan, J. (2007), 'Enjoy the joy of copulas: withh a package copula', *Journal of Statistical Software* **21**(1), 1–21.