AStA Advances in Statistical Analysis
https://doi.org/10.1007/s10182-021-00405-9

ORIGINAL PAPER

Inducing a desired value of correlation between two point-scale variables: a two-step procedure using copulas

Alessandro Barbiero[1]

Received: 17 November 2019 / Accepted: 1 May 2021
© The Author(s) 2021

Abstract

Focusing on point-scale random variables, i.e. variables whose support consists of the first m positive integers, we discuss how to build a joint distribution with pre-specified marginal distributions and Pearson's correlation ρ. After recalling how the desired value ρ is not free to vary between -1 and $+1$, but generally ranges a narrower interval, whose bounds depend on the two marginal distributions, we devise a procedure that first identifies a class of joint distributions, based on a parametric family of copulas, having the desired margins, and then adjusts the copula parameter in order to match the desired correlation. The proposed methodology addresses a need which often arises when assessing the performance and robustness of some new statistical technique, i.e. trying to build a huge number of replicates of a given dataset, which satisfy—on average—some of its features (for example, the empirical marginal distributions and the pairwise linear correlations). The proposal shows several advantages, such as—among others—allowing for dependence structures other than the Gaussian and being able to accommodate the copula parameter up to an assigned level of precision for ρ with a very small computational cost. Based on this procedure, we also suggest a two-step estimation technique for copula-based bivariate discrete distributions, which can be used as an alternative to full and two-step maximum likelihood estimation. Numerical illustration and empirical evidence are provided through some examples and a Monte Carlo simulation study, involving the CUB distribution and three different copulas; an application to real data is also discussed.

Keywords Attainable correlations · Copula · Fréchet bounds · Latent model · Monte Carlo simulation · Ordinal variables

✉ Alessandro Barbiero
alessandro.barbiero@unimi.it

[1] Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan, Italy

Published online: 29 May 2021

1 Introduction

Datasets arising in the social sciences often contain ordinal variables. Sometimes they are genuine ordered assessments (judgements, preferences, degree of liking of a product or adhesion to a sentence, etc.), whereas in other circumstances they are discretized or categorized for convenience (age of people in classes, education achievement, levels of blood pressure, etc.) The former situation often arises when a survey is administered to a group of people being studied, e.g. questionnaires submitted by a company to their customers with the aim of assessing their level of satisfaction towards a product or service the company has provided. Respondents choose a qualitative assessment on a graduated sequence of verbal definitions (for instance, "extremely dissatisfied", "very dissatisfied" , … "very satisfied", "extremely satisfied"), also known as "Likert scale", which can be coded as integers numbers $(1, 2 \ldots, m)$ just for convenience: this amounts to assuming that the categories are evenly spaced (Iannario and Piccolo 2012). There are several statistical models and techniques that can be employed for handling multivariate ordinal data without trying to quantify their ordered categories. (The review by Liu and Agresti (2005) and the later textbook of Agresti (2010) give a thorough treatment.) Among them, correlation models and association models both study departures from independence in contingency tables and involve the assignment of scores to the categories of the row and column variables in order to maximize the relevant measure of relationship (the correlation coefficient in the correlation models or the measure of intrinsic association in association models, see Faust and Wasserman 1993). We also remind nonlinear principal component analysis (NLPCA), which is a special case of a multivariate reduction technique named homogeneity analysis and which can be usefully applied in customer satisfaction surveys (Ferrari and Manzi 2010) for mapping the observed ordinal variables into a one-dimensional (or, more generally, lower-dimensional) quantitative variable. Neither the weights of the original variables nor the differences between their categories (this is the distinction with standard principal component analysis) are assumed a priori.

However, substituting the ordered categories with the corresponding integer numbers, though representing just an arbitrary assumption, is still quite a common and accepted practice, which leads to further multivariate statistical analyses handling them as (correlated) discrete variables (Norman 2010; Carifio and Perla 2008).

Now, one may be interested in building and simulating a multivariate random vector whose univariate components are point-scale variables. In fact, describing a real phenomenon by creating mirror images and imperfect proxies of the (partially) unknown underlying population in a repeated manner allows researchers to study the performance of their statistical methods through simulated data replicates that mimic the real data characteristics of interest in any given setting (Demirtas and Yavuz 2015; Demirtas and Vardar-Acar 2017). This is often necessary since exact analytic results are seldom available for finite sample sizes, and thus, simulation is required to assess the reliability, validity, and plausibility

of inferential techniques and to evaluate to which extent they are robust to deviations from statistical assumptions. However, rather than completely detailing a joint distribution for modelling the phenomena under study, it is often more convenient and realistic to specify only the marginal distributions and pairwise correlations, which are very easy to interpret and whose sample analogues can be easily computed on the dataset at hand one wants to "reproduce". In Lee (1997), some methods are described for generating random vectors of categorical or ordinal variables with specified marginal distributions and degrees of association between variables. For ordinal variables, a common index for measuring association is Goodman and Kruskal's γ coefficient (Kruskal and Goodman 1954; Ruiz and Hüllermeier 2012), ranging between -1 and $+1$, with zero corresponding to independence. A first proposal is based upon using convex combination of joint distribution with extremal values of γ (extremal tables); a second one relies on threshold arguments and involves Archimedean copulas. In this paper, we suggest a sort of modification of this latter method to correlated point-scale variables.

In the following, we will limit our analysis to the bivariate case, which is by far easier to deal with, but whose results, with some caution, can be generalized to the multivariate context. We consider two point-scale random variables (rvs), X_1 and X_2, defined over the support spaces $\mathcal{X}_1 = \{1, 2, \ldots, m_1\}$ and $\mathcal{X}_2 = \{1, 2, \ldots, m_2\}$, respectively, with probability mass functions $p_1(i) = p_{i\cdot} = P(X_1 = i), i = 1, \ldots, m_1$, and $p_2(j) = p_{\cdot j} = P(X_2 = j), j = 1, \ldots, m_2$. We want to determine *some* bivariate probability mass function $p_{ij} = p(i,j) = P(X_1 = i, X_2 = j), i = 1, \ldots, m_1; j = 1, \ldots, m_2$ such that its margins are p_1 and p_2 and the correlation ρ_{X_1,X_2} is equal to an assigned ρ. In order to give an answer to this question, we have first to recall two properties of Pearson's correlation, which apply to both the continuous and, to even a larger extent, the discrete case; this is the topic of Sect. 2. In Sect. 3, we first state the problem of finding a joint probability function with assigned margins and correlation in general terms; then, we focus on a particular class of joint distributions, recalling how to build copula-based bivariate discrete distributions; finally, we describe the proposed procedure for inducing a desired value of correlation between two point-scale variables. Section 4 illustrates an application to CUB distributions. Section 5 recalls inferential procedures for dependent rvs and based on the algorithm of Sect. 3 devises a sort of moment method for estimating the dependence parameter of the copula-based bivariate distribution. Section 6 describes a Monte Carlo simulation study whose aim is to comparatively assess the statistical performances of the new inferential method and the existing ones based on maximum likelihood. Section 7 provides an application to a real data set. In the concluding section, some final remarks are provided.

2 Attainable correlations between two random variables

Pearson's linear correlation is by far the most popular measure of correlation between two quantitative variables.

It is often employed to measure correlation also between Likert scale variables, which arises some criticism.

Despite the many useful properties it enjoys, it also reveals some disadvantages. A first drawback of Pearson's correlation is that given two marginal cumulative distribution functions (cdfs) F_1 and F_2 and a correlation value $\rho \in [-1, +1]$, it is not always possible to construct a joint distribution F with margins F_1 and F_2, whose correlation is equal to the assigned ρ. This is an issue that is often underrated if not neglected by researchers (Leonov and Qaqish 2020). We can state this result (often reported as "attainable correlations", see McNeil et al. 2005, pp. 204-205) in the following way. Let (X_1, X_2) be a random vector with marginal cdfs F_1 and F_2 and an unspecified joint cdf; assume also that $\text{Var}(X_1) > 0$ and $\text{Var}(X_2) > 0$. The following statements hold:

1. The attainable correlations form a closed interval $[\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} < 0 < \rho_{\max}$.
2. The minimum correlation $\rho = \rho_{\min}$ is attained if and only if X_1 and X_2 are countermonotonic. The maximum correlation $\rho = \rho_{\max}$ is attained if and only if X_1 and X_2 are comonotonic.
3. $\rho_{\min} = -1$ if and only if X_1 and $-X_2$ are of the same type, and $\rho_{\max} = 1$ if and only if X_1 and X_2 are of the same type.

For point-scale rvs X_1 and X_2, it is then clear that the maximum correlation is $+1$ if and only if they are identically distributed: $p_1(i) = p_2(i)$, $\forall i = 1 \dots, m$, with $m = m_1 = m_2$, whereas the minimum correlation is -1 if and only if $m_1 = m_2 = m$ and $p_1(i) = p_2(m - i + 1)$, $\forall i = 1, \dots, m$. In the general case, the values $\rho_{\min}$ and $\rho_{\max}$ can be computed by building the cograduation and countergraduation tables (see Salvemini 1939 and Ferrari and Barbiero 2012 for an example of calculation).

A second result about Pearson's correlation can be resumed as follows: given two margins F_1 and F_2 and a feasible linear correlation ρ (i.e. a value falling within the interval $[\rho_{\min}, \rho_{\max}]$), the joint distribution F having margins F_1 and F_2 and correlation ρ is not unique. In other terms, the marginal distributions and pairwise correlation of a bivariate rv do not univocally determine its joint distribution. Even if this second fallacy may represent a limit from one side, on the other side represents a form of flexibility, since it means that given two point-scale distributions and a consistent value of ρ, there are generally several (possibly, infinite) different ways to join them into a bivariate distribution with that value of correlation, as we will see in the next two sections.

3 A procedure for inducing a desired value of correlation between two point-scale random variables with assigned marginal distributions

We will now state the problem object of this work in general terms; then, resorting to copulas, we will reformulate it in a more specific context. Somehow, we will split the original problem into two sequential sub-problems: (i) finding a

family of joint distributions with the assigned margins, (ii) finding within this family a distribution with the desired value of correlation.

3.1 Statement of the problem

The problem of finding a bivariate point-scale distribution with assigned marginal distributions and Pearson's correlation can be laid out as follows. We have to find the $m_1 \times m_2$ probabilities p_{ij}, $0 \le p_{ij} \le 1$, defining the joint pmf of the rv (X_1, X_2), which satisfy the following system of equalities:

$$\begin{cases} \sum_{j=1}^{m_2} p_{ij} = p_{i\cdot}, \quad i = 1, \dots, m_1 \\ \sum_{i=1}^{m_1} p_{ij} = p_{\cdot j}, \quad j = 1, \dots, m_2 \\ \rho_{x_1 x_2} = (\mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2))(\mathrm{Var}(X_1)\mathrm{Var}(X_2))^{-1/2} \end{cases} \tag{1}$$

with $\mu_1 = \mathbb{E}(X_1) = \sum_{i=1}^{m_1} i p_{i\cdot}$, $\mathrm{Var}(X_1) = \sum_{i=1}^{m_1} (i - \mu_1)^2 p_{i\cdot}$ (analogous results hold for X_2), and $\mathbb{E}(X_1 X_2) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} ijp_{ij}$. The first two equalities correspond to the request of matching the assigned marginal distributions and the latter to the assigned correlation. The total number of equality constraints is $m_1 + m_2$ ($m_1 + m_2 - 1$ actual constraints on the two margins, plus one on Pearson correlation).

If, for example, $m_1 = m_2 = 2$ (i.e. if X_1 and X_2 are shifted Bernoulli rvs), then we have a system of 4 equations in 4 variables, which yields a unique solution for the p_{ij}. In this case, in fact, one can easily prove that if the assigned ρ falls within the bounds $\rho_{\min}$ and $\rho_{\max}$, equal to

$$\rho_{\max} = [\min(1 - p_{1\cdot}, 1 - p_{\cdot 1}) - (1 - p_{1\cdot})(1 - p_{\cdot 1})] / \sqrt{p_{1\cdot}(1 - p_{1\cdot}) p_{\cdot 1}(1 - p_{\cdot 1})}$$
$$\rho_{\min} = [\max(0, 1 - p_{\cdot 1} - p_{1\cdot}) - (1 - p_{1\cdot})(1 - p_{\cdot 1})] / \sqrt{p_{1\cdot}(1 - p_{1\cdot}) p_{\cdot 1}(1 - p_{\cdot 1})}$$

the probabilities satisfying the system (1) are

$$\begin{cases} p_{22} &= (1 - p_{1\cdot})(1 - p_{\cdot 1}) + \rho \sqrt{p_{1\cdot}(1 - p_{1\cdot}) p_{\cdot 1}(1 - p_{\cdot 1})} \\ p_{12} &= 1 - p_{\cdot 1} - p_{22} \\ p_{21} &= 1 - p_{1\cdot} - p_{22} \\ p_{11} &= p_{1\cdot} - p_{12} \end{cases}$$

For higher values of m_1 (and m_2) the solution is not unique; generally, there are infinite solutions (i.e. bivariate distributions) satisfying system (1)—given that the bivariate correlation bounds are respected (i.e. $\rho_{\min} \le \rho \le \rho_{\max}$). We illustrate it through the following example.

Example 1 (Computation of joint probabilities for a 2×3 contingency table). Let us consider a rv (X_1, X_2), whose marginals distributions are $p_{1\cdot} = 1/3$, $p_{2\cdot} = 2/3$; $p_{\cdot 1} = 1/4$, $p_{\cdot 2} = 1/2$, $p_{\cdot 3} = 1/4$, with support $\{1, 2\}$ and $\{1, 2, 3\}$, respectively. Marginal expected values and variances are: $\mathbb{E}(X_1) = 5/3$, $\mathbb{E}(X_2) = 2$; $\mathrm{Var}(X_1) = 2/9$, $\mathrm{Var}(X_2) = 1/2$. The bounds for Pearson's correlation, based on countergraduation

Table 1 Cograduation and countergraduation tables corresponding to the marginal distributions of Example 1

(i, j)	1	2	3	Total
(a) Cograduation table				
1	1/4	1/12	0	1/3
2	0	5/12	1/4	2/3
Total	1/4	1/2	1/4	1
(b) Countergraduation table				
1	0	1/12	1/4	1/3
2	1/4	5/12	0	2/3
Total	1/4	1/2	1/4	1

and cograduation tables (see Table 1a and b), are $\rho_{\min} = -3/4$ and $\rho_{\max} = 3/4$. Let the target correlation value be $\rho = 1/3$. Then, system (1) consists of the following 6 equations in 6 variables:

$$\begin{cases} p_{11} + p_{12} + p_{13} = 1/3 \\ p_{21} + p_{22} + p_{23} = 2/3 \\ p_{11} + p_{21} = 1/4 \\ p_{12} + p_{22} = 1/2 \\ p_{13} + p_{23} = 1/4 \\ 3(p_{11} + 2p_{12} + 3p_{13} + 2p_{21} + 4p_{22} + 6p_{23} - 10/3) = 1/3 \end{cases}$$

Solving the system above, imposing that each p_{ij} must be non-negative, we obtain that a valid joint pmf respecting the assigned margins and the correlation coefficient is derived as

$$\begin{cases} p_{12} &= 4/9 - 2p_{11} \\ p_{13} &= p_{11} - 1/9 \\ p_{21} &= 1/4 - p_{11} \\ p_{22} &= 2p_{11} + 1/18 \\ p_{23} &= 13/36 - p_{11} \end{cases}$$

Table 2 Examples of joint distributions with the same assigned margins and the same correlation coefficient $\rho = 1/3$; see Example 1

(i, j)	1	2	3	Total
(a)$p_{11} = 1/6$				
1	1/6	1/9	1/18	1/3
2	1/12	7/18	7/36	2/3
Total	1/4	1/2	1/4	1
(b) $p_{11} = 1/9$				
1	1/9	2/9	0	1/3
2	5/36	5/18	1/4	2/3
Total	1/4	1/2	1/4	1

with p_{11} taking any value contained in the interval [1/9, 2/9]; for example, if we set $p_{11} = 1/6$, we obtain the contingency table displayed in Table 2a; letting $p_{11} = 1/9$, the result is Table 2b.

Instead of considering the whole set of feasible joint probabilities p_{ij} obtained by solving system (1), one can restrict the analysis to a particular subset of these solutions satisfying the first two constraints on margins. For this aim, we will now recall the concept of copula and copula-based bivariate discrete distributions.

3.2 Generating bivariate discrete distributions having the pre-specified margins

Using copulas represents a straightforward solution for easily constructing a multivariate distribution respecting the assigned margins. A d-dimensional copula C is a joint cdf in $[0,1]^d$ with standard uniform margins U_j, $j = 1 \dots, d$:

$$C(u_1, \dots, u_d) := P(U_1 \leq u_1, \dots, U_d \leq u_d), \quad (u_1, \dots, u_d) \in [0,1]^d.$$

The importance of copulas in studying multivariate cdfs is summarized by the Sklar's theorem (McNeil et al. 2005), whose version for $d = 2$ states that if F_1 and F_2 are the cdfs of the rvs X_1 and X_2, the function

$$F(x,y) = C(F_1(x), F_2(y)), \quad x, y \in \mathbb{R}, \tag{2}$$

defines a valid joint cdf, whose margins are exactly F_1 and F_2. This result keeps holding if X_1 and X_2 are point-scale rvs; in this case, the joint pmf can be derived from (2) as:

$$p(i,j) = F(i,j) - F(i-1,j) - F(i,j-1) + F(i-1,j-1), \tag{3}$$

for $i = 1, \dots, m_1; j = 1, \dots, m_2$. It is worth noting that given a joint cdf F with margins F_1 and F_2, Sklar's theorem also states that there exists a copula C such that F can be written as in (2). This copula is unique if F_1 and F_2 are continuous; on the contrary, uniqueness is not guaranteed if they are discrete.

There exists a multitude of copulas, which—as happens for joint cdfs—usually depend on some parameter θ; we will now review three well-known parametric copula families.

3.2.1 The Gauss copula

The d-variate Gauss copula is the copula that can be extracted from a d-variate normal vector $\boldsymbol{Y}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ and is exactly the same as the copula of $\boldsymbol{X} \sim \mathcal{N}_d(\mathbf{0}, P)$, where P is the correlation matrix of $\boldsymbol{Y}$. In two dimensions, it can be expressed, for $\rho_{Ga} \neq \pm 1$, as:

$$C^{Ga}(u_1, u_2; \rho_{Ga}) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho_{Ga}^2}} e^{-\frac{s_1^2 - 2\rho_{Ga} s_1 s_2 + s_2^2}{2(1-\rho_{Ga}^2)}} \mathrm{d}s_1 \mathrm{d}s_2.$$

Independence, comonotonicity, and countermonotonicity copulas are special cases of the bivariate Gauss copula (for $\rho_{Ga} = 0$, $\rho_{Ga} \to +1$, and $\rho_{Ga} \to -1$, respectively).

3.2.2 The Frank copula

The one-parameter bivariate Frank copula is defined as

$$C^F(u_1, u_2; \kappa) = -\frac{1}{\kappa} \ln \left[1 + \frac{(e^{-\kappa u_1} - 1)(e^{-\kappa u_2} - 1)}{e^{-\kappa} - 1} \right],$$

with $\kappa \neq 0$. For $\kappa \to 0$, we have that the Frank copula reduces to the independence copula; for $\kappa \to +\infty$, it tends to the comonotonicity copula; for $\kappa \to -\infty$, it tends to the countermonotonicity copula.

3.2.3 The Plackett copula

The one-parameter bivariate Plackett copula is defined as

$$C^P(u_1, u_2; \theta) = \frac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{[1 + (\theta - 1)(u_1 + u_2)]^2 - 4\theta(\theta - 1)u_1 u_2}}{2(\theta - 1)},$$

with $\theta \in (0, +\infty) \setminus \{1\}$. When $\theta \to 1$, it reduces to the independence copula, whereas for $\theta \to 0$ it tends to the countermonotonicity copula and for $\theta \to \infty$ to the comonotonicity copula.

3.3 An algorithm for inducing any feasible value of correlation within a parametric copula-based family of distributions

In order to induce any feasible value of correlation between the two discrete margins of the distribution (2), we have further to impose that the copula $C(\cdot;\theta)$ is able to encompass the entire range of dependence, from perfect negative dependence (which leads to the linear correlation $\rho_{\min}$) to perfect positive dependence ($\rho_{\max}$). Copulas enjoying this property are named "comprehensive"; the three copulas recalled in the previous section are all comprehensive.

Once the marginal distributions of X_1 and X_2 are assigned, and the parametric copula $C(\cdot;\theta)$ has been selected, their correlation coefficient ρ_{X_1,X_2} will depend only on the copula parameter $\theta \in [\theta_{\min}, \theta_{\max}]$; this relationship may be written in an analytical or numerical form, say $\rho_{X_1,X_2} = g(\theta|F_1, F_2)$. Since the function g is not usually analytically invertible, inducing a desired value of correlation ρ between two point-scale variables, falling in $[\rho_{\min}, \rho_{\max}]$, by setting an appropriate value of θ, is a task that can be generally done only numerically, by finding the (unique) root of the equation $g(\theta) - \rho = 0$. If $g(\theta)$ is a monotone increasing function of the copula parameter, it can be implemented by resorting to the following iterative procedure (see a similar proposal for the Gauss copula in Ferrari and Barbiero 2012; Barbiero and Ferrari 2017; and an early extension to other copulas in Barbiero 2018):

1. Set $\theta^{(0)} = \theta^{\Pi}$ (with θ^{Π} being the value of θ for which the copula C reduces to the independence copula); $\rho^{(0)} = 0$.
2. Set $t \leftarrow 1$ and $\theta = \theta^{(t)}$, with $\theta^{(t)}$ some value strictly greater (smaller) than $\theta^{(0)}$ if $\rho > (<)0$
3. Compute $F(i,j;\theta^{(t)})$ using (2)
4. Compute $p(i,j;\theta^{(t)})$ using (3)
5. Compute $\rho^{(t)}$
6. If $|\rho^{(t)} - \rho| < \epsilon$ stop; else set $t \leftarrow t+1$, $\theta^{(t)} \leftarrow \min(\theta_{\max}, \theta^{(t-1)} + m^{(t)}(\rho - \rho^{(t-1)}))$ if $\rho > 0$, or $\theta^{(t)} \leftarrow \max(\theta_{\min}, \theta^{(t-1)} + m^{(t)}(\rho - \rho^{(t-1)}))$ if $\rho < 0$, with $m^{(t)} = \dfrac{\theta^{(t-1)} - \theta^{(t-2)}}{\rho^{(t-1)} - \rho^{(t-2)}}$; go back to 3.

The iterative process at the basis of the algorithm is quite clear (see Fig. 1): one starts from two points in the (θ, ρ) Cartesian diagram: $A = (\theta^{(0)}, \rho^{(0)})$ and $B = (\theta^{(1)}, \rho^{(1)})$, where $\theta^{(1)}$ can be chosen arbitrarily, respecting the unique condition that the resulting $\rho^{(1)}$ has the same sign as the target ρ. From these two points, one derives the next value of θ, $\theta^{(2)}$ (corresponding to the abscissa of point C), by linear interpolation, considering the slope $m^{(2)}$, associated with the line passing through them, respecting the lower or upper bounds of θ (this is why the min and max operators appear in the recursive formulas for $\theta^{(t)}$ of step 6); the procedure then continues, computing the actual value $\rho^{(2)}$ (ordinate of point D), and then iteratively updating $\theta^{(t)}$ (and computing $\rho^{(t)}$) by taking into account just the last two points for determining the updated $m^{(t)}$.

The above heuristic algorithm makes sense if g is a monotone increasing function, which is often the case: for the Gauss, Frank, and Plackett copulas, the linear correlation is an increasing function of the dependence parameter θ, keeping fixed the two marginal distributions.

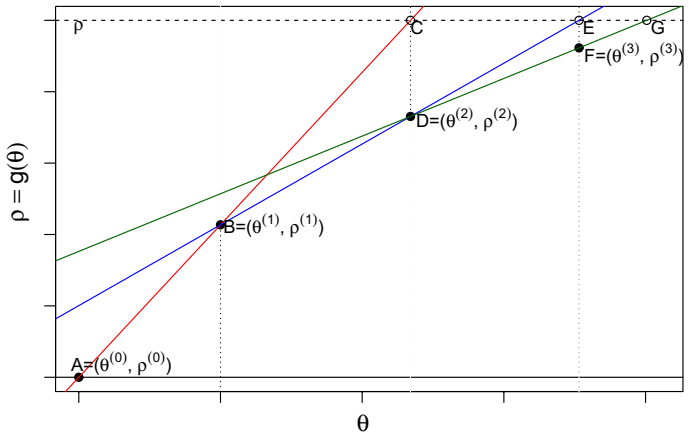

Fig. 1 Graphical representation of the algorithm proposed for inducing a feasible value of correlation between point-scale rvs. The abscissas of black-filled points (A, B, D, F, ⋯) represent the succession of the values $\theta^{(t)}$ of the copula parameter, converging to the solution of the problem (1)

In fact, let us recall that we say that the joint cdf $F(x_1, x_2;\theta)$, with fixed margins F_1 and F_2, is "increasing in concordance" as θ increases if, for any $\theta_2 > \theta_1$,

$$F(x_1, x_2;\theta_1) \leq F(x_1, x_2;\theta_2) \quad \forall x_1, x_2 \in \mathbb{R}.$$

Then, it follows (see, for example, Scarsini and Shaked 1996)

$$\rho_{X_1,X_2}(\theta_1) \leq \rho_{X_1,X_2}(\theta_2).$$

Since the Gauss copula, the Frank copula, and the Plackett copula are all increasing in concordance with respect to their parameter, i.e.

$$C(u_1, u_2;\theta_1) \leq C(u_1, u_2;\theta_2),$$

(Joe 2014) and the same holds for the joint cdf $F(x_1, x_2;\theta) = C(F_1(x_1), F_2(x_2);\theta)$, then we can claim that ρ_{X_1,X_2} is increasing in θ.

A very particular case is represented by the Gauss copula. A known theoretical result, which goes under the name of Lancaster theorem (Lancaster 1957, p.290), and later reported for example in Cario and Nelson (1997), allows us to claim that the correlation between the discrete rvs X_1 and X_2 has the same sign of and in absolute value it is not greater than the Gauss copula correlation: $\text{sgn}(\rho_{X_1,X_2}) = \text{sgn}(\rho_{Ga})$ and $|\rho_{X_1,X_2}| \leq |\rho_{Ga}|$. Therefore, a reasonable value for the starting value $\theta^{(1)} := \rho_{Ga}^{(1)}$ is the value of the target correlation itself.

The advantage of the proposed algorithm stands in the four following (connected) features: (i) the flexibility in the choice of the underlying copula, which can be different from the Gaussian and is just required to span the entire dependence range, (ii) the capacity of finding the appropriate value of θ without making use of any sample from the two marginal distributions, thus avoiding introducing sampling errors, (iii) the possibility of controlling a priori the error ϵ (absolute difference between target and actual value of ρ_{X_1,X_2})—setting ϵ equal to 10^{-7} generally allows to recover θ in a few steps, and (iv) the absence of inner potentially time-consuming optimization or root finding routines.

Existing procedures for solving the same (or a similar) problem are available in the literature, but do not enjoy all the features above mentioned. For example, the proposal by Demirtas (2006) is based on simulating binary data whose marginals are derived collapsing the prespecified marginals of ordinal variables. The correlation matrix of the binary variables is obtained by an iterative process in order to match the target correlation matrix for ordinal data, which requires the generation of a "huge" bivariate sample of binary data.

Other proposals by Madsen and Dalthorp (2007), Ferrari and Barbiero (2012), or Xiao (2017) for simulating multivariate ordinal/discrete variables with assigned margins and correlations exclusively address the dependence structure induced by the Gauss copula.

Lee and Kaplan (2018) proposed two procedures based on the principles of maximum entropy and minimum cross-entropy to simulate multivariate ordinal variables with assigned values of marginal skewness and kurtosis; they rely on the multivariate normal distribution as a latent variable. Foldnes and Olsson (2016) proposed a simulation technique for nonnormal data with pre-specified skewness, kurtosis,

and covariance matrix, by using linear combinations of independent generator (IG) variables; its most important feature is that the resulting copula is not Gaussian. In Nelsen (1987), using convex linear combinations of the pmfs for the discrete Fréchet boundary distributions (i.e. those corresponding to comotonicity and countermonotonicity) and the pmf for independent rvs, the author constructs bivariate pmfs for dependent discrete rvs with arbitrary marginals and any correlation between the theoretical minimum and maximum. A similar rationale has been later used by Demirtas and Vardar-Acar (2017) for devising an algorithm for inducing any desired Pearson or Spearman correlation to independent bivariate data whose marginals can be of any distributional type and nature. The algorithm we proposed, though limited to point-scale rvs, is much more flexible as it allows a much broader choice of dependence structures, whereas the latter two procedures employ a convex combination of bivariate comonotonicity and countermonotonicity copulas. (An analogous way was followed by Lee (1997) for the first method.)

Obviously, in the simplest case of a bivariate shifted Bernoulli ($m_1 = m_2 = 2$), the proposed algorithm recovers (numerically) the same unique bivariate distribution yielded (analytically) by system (1), whatever copula is selected (provided it spans the entire dependence spectrum, i.e. it is comprehensive). For the case of dependent Bernoulli rvs, see also the example presented in McNeil et al. 2005, p.188.

We remark that even if we mentioned three well-known comprehensive copulas (Gauss, Frank, and Plackett), which are all exchangeable and radially symmetric (see, for example, McNeil et al. (2005), chapter 5), exchangeability and radial symmetry are not necessary conditions for the algorithm to work; thus, the proposed procedure is able to deal with asymmetrical dependence, which often occurs in many fields, especially in finance. The comprehensive property is instead required if one wants to span the entire range of feasible linear correlations between the two point-scale rvs. If one uses the Gumbel or the Clayton copula (both belonging to the broad class of Archimedean copulas) to induce correlation between the rvs, since these two copulas can only model positive dependence through their scalar parameter θ, it descends that only positive (or at most null) values of linear correlation can be induced and then assigned. A useful reference is Table 4.1 in Nelsen (2006), where some important one-parameter families of Archimedean copulas are listed along with some special and limiting cases; from there, it is thus possible to distinguish and select comprehensive copulas.

The algorithm presented in this section is naturally conceived for one-parameter copulas, but it can be extended to p-parameter copulas, $p \geq 2$ (just think of Student's t copula); in this case, $p - 1$ higher-order correlations or co-moments need to be assigned along with the usual linear correlation in order to calibrate all the parameters.

3.4 Extension to multivariate context

The extension of the proposed procedure to dimension $d > 2$ —finding a joint distributions with d assigned margins and $d(d-1)/2$ distinct pairwise correlations—is not straightforward at all, but this is due to theoretical rather

than computational reasons. To explain it, we will refer to a counterexample described in Bergsma and Rudas (2002) and Chaganty and Joe (2006). Let X, Y and Z be three correlated binary rvs, each with support $\{0, 1\}$, such that $P(X = 1) = P(Y = 1) = P(Z = 1) = 0.5$ (nothing changes if we select $\{1, 2\}$ as common support). Based on the bivariate correlation bounds (see point 3 in Sect. 2) the three correlation coefficients ρ_{XY} , ρ_{XZ}, ρ_{YZ} can lie in the interval $[-1, +1]$. However, if we choose $\rho_{XY} = \rho_{YZ} = 0.4$ and $\rho_{XZ} = -0.4$, then a trivariate distribution for (X, Y, Z) does not exist. So, a first type of problem is related to the feasibility of the (assigned) correlation matrix P for the discrete d-variate random vector: even if all the bivariate correlation bounds are respected and even if the matrix P reckoning all the pairwise correlations is a valid correlation matrix (a symmetric matrix with all ones on the main diagonal, which is positive semidefinite), nevertheless it may be impossible to construct a random vector with the d assigned margins and correlation matrix P. A second type of problem is related to the nearly lack of copulas able to calibrate—through their parameters—the values of the resulting pairwise correlation coefficients. The typical generalizations of the Frank and Plackett copulas, which we discussed in Sect. 3.2, are still characterized by a unique scalar parameter. Assigning arbitrary—though feasible—values to the pairwise correlations of the discrete random vector, generally would lead to no solution in terms of the dependence parameter. A way to overcome this issue is resorting to a copula whose number of parameters is at least equal to the number of distinct pairwise correlations: an obvious candidate is the Gauss copula; a richer option is represented by the t copula. Alternatively, one can resort to pair copula construction through vines, which are graphical models that represent high-dimensional distributions and can model a rich variety of dependencies. Another limitation of the algorithm of Sect. 3.3 is that it only handles rvs with finite supports: extensions to count variables need to seek for an accurate approximation of the correlation coefficient at step 5, possibly entailing a truncation of the support of the joint distribution of steps 3 and 4, in order to compute the correlation coefficient through a double finite summation.

3.5 Pseudo-random simulation

Simulating samples from a bivariate rv with assigned point-scale margins and correlation, built according the procedure described in Sects. 3.2 and 3.3, is straightforward. One can resort to the following general algorithm for meta-copula distributions:

1. Simulate a random sample (u_1, u_2) from the copula $C(u_1, u_2;\theta)$, where θ is the value of the copula parameter recovered through the algorithm of Section 3.3;
2. Set $x_1 = F_1^{-1}(u_1)$ and $x_2 = F_2^{-1}(u_2)$, where F_1^{-1} and F_2^{-1} are the generalized inverse functions of F_1 and F_2, respectively;
3. (x_1, x_2) is a random sample from the target bivariate distribution, with copula $C(u_1, u_2;\theta)$ and margins F_1 and F_2.

Alternatively, since both X_1 and X_2 have a finite support space, one can resort to the "inversion algorithm" described in (Devroye (1986), pp.85,559) and used in Lee (1997): in this case, one directly considers the joint probability mass function $p(i, j)$ of Eq.(3) and proceeds as follows:

1. Set $N = m_1 \times m_2$; let π_t, $t = 1, \ldots, N$ be the joint probabilities p_{ij} arranged in descending order;
2. Let y_t be the corresponding possible values of $Y = (X_1, X_2)$, arranged similarly;
3. Define $z_0, z_1, \ldots, z_N$ in the following way:

$$z_0 = 0$$
$$z_t = z_{t-1} + \pi_t, \quad t = 1, \ldots, N$$

4. Simulate a random number u from a standard uniform rv;
5. Return y_t, where $z_{t-1} < u \leq z_t$.

4 Application to CUB random variables

A CUB rv X is defined as the mixture, with weights π and $1 - \pi$, of a shifted binomial with parameters m and ξ and a discrete uniform distribution over the support $\{1, 2, \ldots, m\}$, for $m > 3$ (Piccolo 2003). Its probability mass function is thus

$$P(X = i) = \pi \binom{m-1}{i-1} \xi^{m-i} (1-\xi)^{i-1} + (1-\pi)\frac{1}{m}, \quad i = 1, 2, \ldots, m,$$

with (π, ξ) a parameter vector with parameter space $\Theta = (0, 1] \times [0, 1]$. This rv is shown to be adapt to model the discrete choice of an individual from an ordinal list of m alternatives, which can be regarded as the result of a complex decision blending both the attractiveness ("feeling") towards an item, expressed by $1 - \xi$, and fuzziness surrounding the final choice ("uncertainty") expressed by $1 - \pi$. The requirement $m > 3$ allows the identifiability of the model as it avoids the cases of a degenerate rv (if $m = 1$), an indeterminate model (if $m = 2$) and a saturated model (if $m = 3$). With only two parameters, the CUB rv is able to model extreme and intermediate modes, positive and negative skewness, flat and peaked behaviours, etc. (see Piccolo 2003; Iannario and Piccolo 2010, for further properties of this stochastic model).

Corduas (2011) proposed using the Plackett copula in order to construct a one-parameter bivariate distribution from CUB margins; this proposal was later investigated by Andreis and Ferrari (2013), also in a multivariate direction. Here, we reprise and extend these attempts of constructing a bivariate CUB rv, by resorting to the results discussed in Section 3.2. Let us suppose that we want to build a bivariate model with margins $X_1 \sim \text{CUB}(m_1 = 5, \pi_1 = 0.4, \xi_1 = 0.8)$ and $X_2 \sim \text{CUB}(m_2 = 5, \pi_2 = 0.7, \xi_2 = 0.3)$; we can find the values of the attainable correlations by using the function `corrcheck` in `GenOrd` (Barbiero and Ferrari 2015a), which returns the values $\rho_{\min} = -0.952003$ and $\rho_{\max} = 0.8640543$. We

proceed and select a desired feasible value of correlation between the two CUB variates, say $\rho = 0.6$.

Afterward, we recover the values of ρ_{Ga} (for the Gauss copula), κ (for the Frank copula), and θ (for the Plackett copula), according to the iterative procedure illustrated in the previous section. By setting $\epsilon = 10^{-7}$, we obtain $\rho_{Ga} = 0.6898959$, $\kappa = 5.453455$, and $\theta = 11.30106$. Table 3 reports the detailed iterations of the algorithm for the three copula-based models. Note that since $\rho > 0$, for the Frank copula we selected a value $\kappa^{(1)}$ larger than zero (tentatively, $\kappa^{(1)} = 1$) and for the

Table 3 Iterations of the algorithm when applied to the example CUB model. Note that for (b) and (c) the $\{\theta^{(t)}\}$ is an increasing sequence

t	$\rho_{Ga}^{(t)}$	$\rho_{X_1,X_2}^{(t)}$
(a) Gauss copula		
0	0	0
1	0.6	0.5216952
2	0.690058	0.6001415
3	0.6898956	0.5999997
4	0.6898959	0.6
(b) Frank copula		
t	$\theta^{(t)}$	$\rho_{X_1,X_2}^{(t)}$
0	0	0
1	1	0.1485864
2	4.038055	0.5001689
3	4.900705	0.5652346
4	5.361629	0.594565
5	5.44704	0.5996245
6	5.453378	0.5999955
7	5.453455	0.6
Plackett copula		
t	$\theta^{(t)}$	$\rho_{X_1,X_2}^{(t)}$
0	1	0
1	2	0.2046548
2	3.931766	0.3839207
3	6.260232	0.4895515
4	8.694896	0.5542904
5	10.41392	0.5863079
6	11.14905	0.5977669
7	11.29231	0.5998727
8	11.30097	0.5999988
9	11.30106	0.6

Plackett copula a value $\theta^{(1)}$ larger than one (tentatively, $\theta^{(1)} = 2$). For the Gauss copula, we set $\theta^{(1)} = \rho = 0.6$.

The three joint pmfs, sharing the same value of linear correlation, are reported in Table 4. It is easy to notice the differences among them. For example, the probability $p_{23} = P(X_1 = 2, X_2 = 3)$ takes the values 0.0922, 0.0948, and 0.1008, across the three joint distributions.

Figure 2 displays the relationship between the copula parameter θ (on the x axis) and the corresponding Pearson's correlation of the example bivariate CUB model with Gauss, Frank, and Plackett copula, constructed by applying steps $3 \div 5$ of the proposed algorithm over a dense grid of uniformly spaced values of θ. The almost linear trend for the Gauss copula can be easily noted (in this case, the copula parameter is itself a correlation coefficient!), whereas for the other two it is obviously (highly) nonlinear, due also to the unbounded domain for κ and θ. From the figure, one can state that ρ_{X_1,X_2} is a monotone concave function of the copula parameter for the Plackett copula, and for the Frank copula only when $\kappa > 0$: this explains the increasing nature of the sequences $\left\{\theta^{(t)}\right\}$ in Tables 3b and 3c.

Table 4 Bivariate distribution with margins $X_1 \sim \text{CUB}(m_1 = 5, \pi_1 = 0.4, \xi_1 = 0.8)$ and $X_2 \sim \text{CUB}(m_2 = 5, \pi_2 = 0.7, \xi_2 = 0.3)$ and $\rho_{x_1 x_2} = 0.6$, obtained based on different copulas

(i, j)	1	2	3	4	5	Total
Gauss copula						
1	0.0553	0.0711	0.0959	0.0551	0.0065	0.2838
2	0.0088	0.0317	0.0922	0.1178	0.0333	0.2838
3	0.0013	0.0077	0.0377	0.0869	0.0479	0.1814
4	0.0002	0.0020	0.0150	0.0566	0.0565	0.1302
5	0.0000	0.0004	0.0045	0.0319	0.0838	0.1206
Total	0.0657	0.1129	0.2452	0.3481	0.2281	1
Frank copula						
1	0.0498	0.0744	0.1042	0.0483	0.0071	0.2838
2	0.0126	0.0297	0.0948	0.1167	0.0300	0.2838
3	0.0022	0.0060	0.0301	0.0916	0.0515	0.1814
4	0.0007	0.0019	0.0108	0.0548	0.0621	0.1302
5	0.0003	0.0009	0.0053	0.0366	0.0775	0.1206
Total	0.0657	0.1129	0.2452	0.3481	0.2281	1
Plackett copula						
1	0.0518	0.0775	0.1001	0.0439	0.0105	0.2838
2	0.0093	0.0251	0.1008	0.1221	0.0266	0.2838
3	0.0025	0.0060	0.0276	0.1004	0.0450	0.1814
4	0.0012	0.0026	0.0105	0.0532	0.0627	0.1302
5	0.0008	0.0018	0.0062	0.0285	0.0833	0.1206
Total	0.0657	0.1129	0.2452	0.3481	0.2281	1

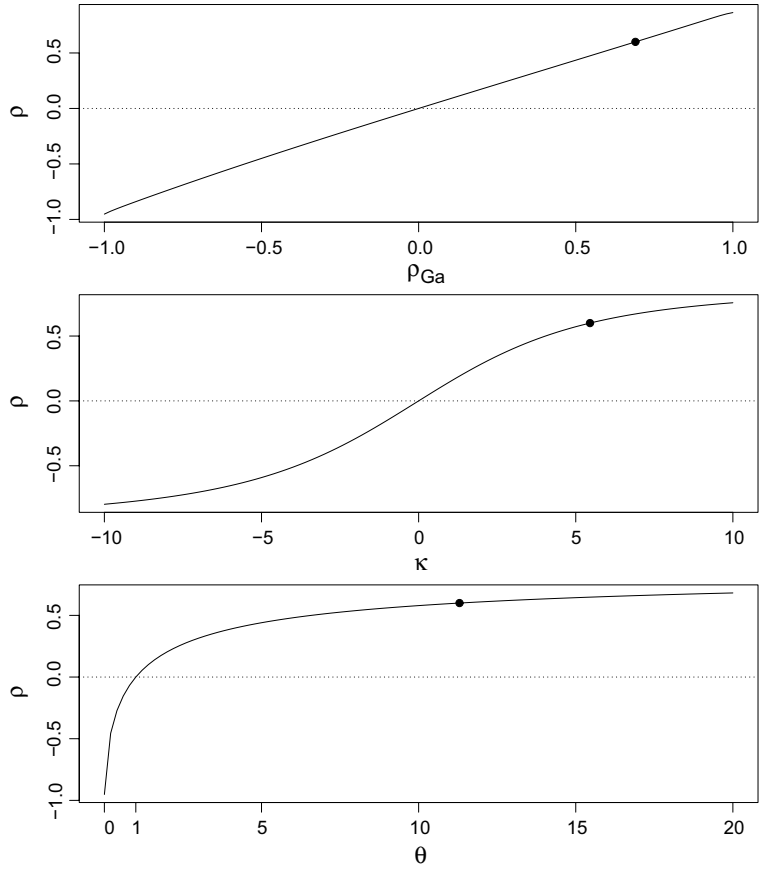

Fig. 2 Relationship between linear correlation and dependence parameter for a bivariate CUB model with Gaussian (top panel), Frank (centre), and Plackett (bottom) copula. Points are drawn in correspondence of $\rho = 0.6$

5 Inferential aspects

If we have a bivariate ordinal sample (x_{1t}, x_{2t}), $t = 1, \dots, n$, that we assume has been drawn from a joint cdf $F(x_1, x_2; \theta, \theta_1, \theta_2) = C(F_1(x_1; \theta_1), F_2(x_2; \theta_2); \theta)$, then parameters' estimation can be carried out through different inferential techniques. First, let us define the log-likelihood function ℓ as

$$\ell(\theta_1, \theta_2, \theta) = \ell(\theta_1, \theta_2, \theta | (x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})) = \sum_{t=1}^{n} \log p(x_{1t}, x_{2t}; \theta_1, \theta_2, \theta)$$
$$= \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \log p(i, j; \theta_1, \theta_2, \theta) \tag{4}$$

with p being the joint pmf of Eq.(3) and n_{ij} the absolute joint frequency of (x_i, x_j), $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$.

We now present three possible ways to perform point estimation of the distribution's parameters: the first one is the customary maximum likelihood method; the second one is a modification thereof, whose use with copula models is however quite consolidated; the third one is directly suggested by the simulation procedure of Sect. 3.3 and can be regarded to as a by-product of the methodology presented in Sect. 3.

5.1 Full maximum likelihood

The most standard estimation technique consists of maximizing ℓ with respect to all the three parameters (or parameter vectors) simultaneously. This task can be usually done just numerically (i.e. no closed-form expressions are available for the parameter estimates), by resorting to customary optimization routines. The resulting maximum likelihood (ML) estimates can be thus derived as

$$(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}) = \arg \max_{(\theta_1, \theta_2, \theta) \in \Theta} \ell(\theta_1, \theta_2, \theta)$$

where Θ is the parameter space.

5.2 Two-step maximum likelihood

This technique aims at reducing the computational burden of the previous one, by splitting the original maximization problem into two subsequent (sets of) maximizations in lower dimensions. In the first step, one estimates θ_1 and θ_2 separately, by resorting to maximum likelihood estimation, as if the two univariate components of the bivariate sample were independent, i.e. maximizing separately their marginal log-likelihood functions:

$$\ell_1(\theta_1) = \sum_{i=1}^{m_1} n_{i\cdot} \cdot \log p_1(i;\theta_1), \quad \ell_2(\theta_2) = \sum_{j=1}^{m_2} n_{\cdot j} \cdot \log p_2(j;\theta_2),$$

with $n_{i\cdot}$ and $n_{\cdot j}$ being the observed marginal frequencies of X_1 and X_2, respectively, thus finding the estimates $\hat{\theta}_1^{TS}$ and $\hat{\theta}_2^{TS}$ (the superscript "TS" standing for "two-step"). Then, one sets $\theta_1 = \hat{\theta}_1^{TS}$ and $\theta_2 = \hat{\theta}_2^{TS}$ in (4) and maximize it with respect to θ, finding $\hat{\theta}^{TS}$. This technique was introduced in a more general context and exhaustively described in Joe and Xu (1996), where it is also named "inference function for margins" (IFM). The authors compared the efficiency of the IFM with the ML by simulation and found that the ratio of the mean square errors of the IFM estimator to the full ML is close to 1. Theoretically, the ML estimator should be asymptotically the most efficient, since it attains the minimum asymptotic variance bound. However, for finite samples, Patton (2006) found that the IFM was often even more efficient than the ML. As a result, IFM is the main estimation method employed in estimating copula models.

5.3 Two-step maximum likelihood + method of moment

This method is directly suggested by the algorithmic procedure of Sect. 3.3. First, one estimates the marginal parameters θ_1 and θ_2 from sample data x_{1i} and x_{2i}, $i = 1, \ldots, n$, independently maximizing ℓ_1 and ℓ_2 with respect to θ_1 and θ_2, as for the previous technique, obtaining $\hat{\theta}_1^{TS}$ and $\hat{\theta}_2^{TS}$. Then, one considers the maximum likelihood estimates of the marginal cdfs, $\hat{F}_j(\cdot) = F_j(\cdot;\hat{\theta}_j^{TS})$, $j = 1, 2$, and obtain the estimate of the dependence parameter θ via the method of moment, by inverting the relationship between the θ and Pearson's correlation: $\hat{\theta}^{TSM} = g^{-1}(\hat{\rho}_{X_1X_2}; \hat{F}_1, \hat{F}_2)$, where $\hat{\rho}_{X_1,X_2}$ is Pearson's sample correlation coefficient and TSM stands for "two-step-moment method". The evaluation of g^{-1} at $\hat{\rho}_{X_1X_2}$, given $\hat{F}_1$ and $\hat{F}_2$, is carried out through the algorithm of Sect. 3.3.

We remark that these three estimation methods are just possible alternatives to be employed for the specific context of copula-based discrete distributions considered in this work. When dealing with copula-based distributions in the continuous case, a straightforward estimation method for the parameter of a specific bivariate parametric copula is the method of moment. It consists in considering a rank correlation between the two rvs (say, Kendall's τ or Spearman's ρ), looking for a theoretical relationship between it and the parameter of the copula, and substituting the empirical value of the rank correlation into this relationship to derive an estimate of the copula parameter. The main advantage of this method is that it does not require any assumption about the marginal distributions, since rank correlations are margin-free, i.e. they depend on the copula only and are not affected by the marginal distributions.

When dealing with factor copulas, another consolidated estimation method is represented by a sort of simulated method of moments (SMM), where the "moments" that are used in estimation are not the usual ones, but functions of rank statistics. The SMM estimator is derived as the minimizer of the distance between data dependence measures and dependence measures obtained through Monte Carlo simulation of the model (Oh and Patton 2017). In the context of hierarchical Archimedean copulas, Okhrin and Tetereva (2017) investigate a clustering estimator based on Kendall's τ by means of Monte Carlo simulations; it is shown to be competitive in terms of statistical properties (bias and variance) and to be computationally advantageous.

However, using these methods would not be convenient in our context, since we know rank correlations lose their nice properties, holding in the continuous set-up, due to the presence of tied values (see, for example, Nešlehová (2007)).

6 Monte Carlo study

The relative performance of the estimators derived through the three methods described in the previous section, expressed in terms of some statistical indicators such as bias or mean-squared error, can be assessed for finite sample size only via Monte Carlo (MC) simulation. Usually, the estimators of the marginal parameters θ_1 and θ_2 derived through the full maximum likelihood technique and the two-step

methods have a very close statistical behaviour; on the contrary, differences are expected to arise among the estimators of the dependence parameter. Here we will examine the joint behaviour and performance of all the parameters' estimators.

For the multivariate case, we recall that the bias of an estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)'$ of a p-dimensional vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is defined as:

$$\text{bias}(\hat{\boldsymbol{\theta}}) := \mathbb{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta};$$

$\hat{\boldsymbol{\theta}}$ is said to be an unbiased estimator of θ if $\mathbb{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ for any $\boldsymbol{\theta} \in \Theta$.

A multivariate generalization of the mean-squared error (MSE) is provided by the MSE matrix (see, for example, Mittelhammer 2013, p.377):

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'].$$

The MSE matrix can be decomposed into variance and bias components, analogous to the scalar case. Specifically, MSE is equal to the sum of the covariance matrix of $\hat{\boldsymbol{\theta}}$ and the outer product of the bias vector of $\hat{\boldsymbol{\theta}}$:

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \text{cov}(\hat{\boldsymbol{\theta}}) + \text{bias}(\hat{\boldsymbol{\theta}})\text{bias}(\hat{\boldsymbol{\theta}})'.$$

The trace of the MSE matrix defines the expected squared distance (ESD) of the vector estimator $\hat{\boldsymbol{\theta}}$ from the vector estimand $\boldsymbol{\theta}$ and is equal to:

$$\begin{aligned} ESD(\hat{\boldsymbol{\theta}}) &:= \text{tr}(\text{MSE}(\hat{\boldsymbol{\theta}})) = \text{tr}(\mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']) = \mathbb{E}(\text{tr}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})')) \\ &= \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})] = \mathbb{E}\left[\sum_{i=1}^{p}(\hat{\theta}_i - \theta_i)^2\right] \\ &= \sum_{i=1}^{p}\mathbb{E}[(\hat{\theta}_i - \theta_i)^2] = \text{tr}(\text{cov}(\hat{\boldsymbol{\theta}})) + \text{bias}(\hat{\boldsymbol{\theta}})'\text{bias}(\hat{\boldsymbol{\theta}}), \end{aligned}$$

where "tr" denotes the trace of a matrix. Being a scalar, the ESD allows direct and easy comparison among different estimators of the same parameter vector: the lower the value of ESD, the better the estimator. Generally, for the estimator vectors corresponding to the three methods of Sect. 5, such quantities cannot be derived analytically, but an approximation can be obtained through the corresponding MC means computed over S simulation runs:

$$\widehat{ESD}_{MC}(\hat{\boldsymbol{\theta}}) = \frac{1}{S}\sum_{t=1}^{S}\text{tr}(\text{cov}(\hat{\boldsymbol{\theta}}_{\boldsymbol{t}})) + \frac{1}{S}\sum_{t=1}^{S}\widehat{\text{bias}}_{MC}(\hat{\boldsymbol{\theta}})'\widehat{\text{bias}}_{MC}(\hat{\boldsymbol{\theta}})$$

with

$$\widehat{\text{bias}}_{MC}(\hat{\boldsymbol{\theta}}) = \frac{1}{S}\sum_{t=1}^{S}\hat{\theta}_t - \boldsymbol{\theta}$$

and $\hat{\theta}_t$ being the value of the vector estimator for run t, $t = 1, \ldots, S$. The larger the value S, the more accurate the approximation.

A MC study is designed to assess the relative statistical performance of the three types of vector estimators for a bivariate CUB model under an array of artificial scenarios, which are realized by varying the dependence structure, the CUB marginal parameters, and the sample size. We point out that this study will not allow general conclusions, but is intended merely to demonstrate how the different inferential methods work and check the potential of the proposal. As possible dependence structures, we evaluate the Gauss, Frank, and Plackett copulas: the first with parameter ρ_{Ga} equal to -0.6 or $+0.6$; the second with parameter $\kappa = -5$ or $\kappa = +5$; the last with parameter $\theta = 1/4$ or $\theta = 4$. As CUB marginal parameters, we consider $m_1 = m_2 = 5$ and $m_1 = m_2 = 7$, combined with the following values for the marginal parameter vector $\boldsymbol{\theta}_M = (\pi_1, \xi_1, \pi_2, \xi_2)'$: $(0.4, 0.8, 0.4, 0.8)'$, $(0.4, 0.8, 0.7, 0.3)'$, and $(0.7, 0.3, 0.7, 0.3)'$. For all the $2 \times 2 \times 3 = 12$ combinations above, the sample sizes 50 and 100 are investigated, for a total of 24 artificial settings for each type of copula.

Note that the two-step procedures (Sects. 5.2 and 5.3) require that both empirical marginal distributions assume at least 4 distinct values in order the MLEs to be computed (Piccolo 2003). To ensure feasibility of these two techniques for any simulation run, when they occurred, we simply discarded the samples that do not respect such condition from the MC simulation and kept fixed at $S = 1000$ the number of feasible samples to draw and to use for the statistical analysis under each artificial setting.

Simulation results reporting the values of $\widehat{ESD}_{MC}$ are displayed in Tables 5a, b, and c. Other summary indices computed over the 1, 000 simulation runs, such as the MC mean of the bias vector, are available on request.

For the models with the Gauss copula, the full MLE performs the best in terms of ESD; for the other two estimators, the values of ESD are very close to each other for any setting, pointing out a slight preference towards the two-step MLE for $n = 100$. There is also a relevant difference in the values of ESD across the settings; as expected, ESD decreases moving from $n = 50$ to $n = 100$, holding the other parameters fixed. The setting with $\boldsymbol{\theta} = (0.7, 0.3, 0.7, 0.3, -0.6)'$ and $n = 100$ minimizes the value of the index for all the three estimators; the setting with $\boldsymbol{\theta} = (0.4, 0.8, 0.4, 0.8, -0.6)'$ maximizes it.

For the models with the Frank copula, the two-step MLE surprisingly overtakes the full MLE for any setting; the two-step moment estimator shows the worst performance, even if for $n = 100$ this difference is attenuated. Significant improvement of all the three estimators occur when moving from $n = 50$ to $n = 100$; the values of distribution's parameters also affect the values of ESD: however, the effect of the model parameters cannot be easily extracted from the results on the examined scenarios.

For the models with the Plackett copula, we have the following interesting result: the MLE is the best performer (smallest $\widehat{ESD}_{MC}$) when $\theta = 1/4$ (corresponding to negative dependence); for the complementary scenarios, corresponding to $\theta = 4$ and then positive correlation, the two-step MLE is the best performer. The two-step moment estimator has a far worse behaviour than its competitors, apart from one scenario, where it is the second best after full MLE; this is especially apparent for $\theta = 4$. Note that in this case, differently from the previous models, the values

Table 5 Simulation results: Monte Carlo expected squared distance $\widehat{ESD}_{MC}$ computed over 1000 runs

ξ_1	π_1	ξ_2	π_2	ρ_{Ga}	n=50			n=100		
$m_1 = 5, m_2 = 5$					ML	TS	TSM	ML	TS	TSM
(a) For the parameter vector $(\xi_1, \pi_1, \xi_2, \pi_2, \rho_{Ga})$ of the bivariate CUB model with Gaussian copula										
0.4	0.8	0.4	0.8	0.6	0.08317	0.08478	0.08600	0.03656	0.03749	0.03800
0.4	0.8	0.4	0.8	−0.6	0.08660	0.08710	0.08764	0.04102	0.04155	0.04174
0.4	0.8	0.7	0.3	0.6	0.07496	0.07683	0.07657	0.03532	0.03591	0.03608
0.4	0.8	0.7	0.3	−0.6	0.07748	0.07985	0.07999	0.03572	0.03646	0.03688
0.7	0.3	0.7	0.3	0.6	0.06977	0.07094	0.07027	0.03540	0.03548	0.03573
0.7	0.3	0.7	0.3	−0.6	0.07254	0.07602	0.07440	0.03451	0.03490	0.03510
$m_1 = 7, m_2 = 7$					ML	TS	TSM	ML	TS	TSM
0.4	0.8	0.4	0.8	0.6	0.06375	0.06603	0.06796	0.02701	0.02831	0.02912
0.4	0.8	0.4	0.8	−0.6	0.06563	0.06758	0.06840	0.02888	0.02956	0.03003
0.4	0.8	0.7	0.3	0.6	0.05925	0.06028	0.05878	0.02559	0.02635	0.02652
0.4	0.8	0.7	0.3	−0.6	0.05916	0.06404	0.06362	0.02584	0.02675	0.02738
0.7	0.3	0.7	0.3	0.6	0.05303	0.05760	0.05690	0.02514	0.02571	0.02622
0.7	0.3	0.7	0.3	−0.6	0.05363	0.06195	0.05610	0.02442	0.02554	0.02586
ξ_1	π_1	ξ_2	π_2	θ	n=50			n=100		
$m_1 = 5, m_2 = 5$					ML	TS	TSM	ML	TS	TSM
(b) For the parameter vector $(\xi_1, \pi_1, \xi_2, \pi_2, \kappa)$ of the bivariate CUB model with Frank copula										
0.4	0.8	0.4	0.8	5	2.0460	1.9698	2.4226	0.8569	0.8413	0.9439
0.4	0.8	0.4	0.8	− 5	2.1142	1.9738	2.3447	0.8481	0.8385	0.9001
0.4	0.8	0.7	0.3	5	2.1527	2.0362	2.5245	0.8452	0.8300	0.9322
0.4	0.8	0.7	0.3	− 5	2.1080	1.9837	2.5502	0.8420	0.8211	0.9526
0.7	0.3	0.7	0.3	5	2.1932	2.0529	2.2892	0.9154	0.8999	1.1116
0.7	0.3	0.7	0.3	− 5	1.9993	1.9323	2.3184	0.9134	0.8860	1.0401
$m_1 = 7, m_2 = 7$					ML	TS	TSM	ML	TS	TSM
0.4	0.8	0.4	0.8	5	1.7089	1.5823	2.1777	0.7414	0.7309	0.8935
0.4	0.8	0.4	0.8	− 5	1.7371	1.6124	1.9682	0.9134	0.8860	1.0401
0.4	0.8	0.7	0.3	5	1.6528	1.6880	2.1652	0.7767	0.7594	0.8755
0.4	0.8	0.7	0.3	− 5	1.6984	1.6921	2.2656	0.7987	0.7826	0.9734
0.7	0.3	0.7	0.3	5	1.6197	1.6875	2.4327	0.7547	0.7497	1.0419
0.7	0.3	0.7	0.3	− 5	1.7158	2.0467	2.5050	0.7887	0.7853	0.9663
ξ_1	π_1	ξ_2	π_2	θ	n=50			n=100		
$m_1 = 5, m_2 = 5$					ML	TS	TSM	ML	TS	TSM
(c) For the parameter vector $(\xi_1, \pi_1, \xi_2, \pi_2, \theta)$ of the bivariate CUB model with Plackett copula										
0.4	0.8	0.4	0.8	4	5.3621	4.8685	7.6299	2.1374	2.0537	2.6393
0.4	0.8	0.4	0.8	0.25	0.0876	0.0881	0.0914	0.0408	0.0418	0.0433
0.4	0.8	0.7	0.3	4	6.9938	6.3370	8.5883	2.2470	2.1459	2.8445
0.4	0.8	0.7	0.3	0.25	0.0805	0.0818	0.0845	0.0383	0.0390	0.0405

Table 5 (continued)

ξ_1	π_1	ξ_2	π_2	θ	n=50			n=100		
$m_1 = 5, m_2 = 5$					ML	TS	TSM	ML	TS	TSM
0.7	0.3	0.7	0.3	4	5.3892	4.8519	6.6360	2.2952	2.2122	2.9717
0.7	0.3	0.7	0.3	0.25	0.0735	0.0751	0.0787	0.0366	0.0366	0.0390
$m_1 = 7, m_2 = 7$					ML	TS	TSM	ML	TS	TSM
0.4	0.8	0.4	0.8	4	4.7900	4.1620	7.6333	1.8533	1.7625	2.5297
0.4	0.8	0.4	0.8	0.25	0.0700	0.0728	0.0776	0.0313	0.0323	0.0341
0.4	0.8	0.7	0.3	4	4.4099	3.7519	6.8071	2.0048	1.8794	2.7205
0.4	0.8	0.7	0.3	0.25	0.0632	0.0684	0.0706	0.0288	0.0298	0.0315
0.7	0.3	0.7	0.3	4	5.1065	4.3642	7.6463	2.1174	2.0060	2.8883
0.7	0.3	0.7	0.3	0.25	0.0580	0.0812	0.0660	0.0265	0.0273	0.0299

ML full maximum likelihood method, *TS* two-step maximum likelihood method; TSM= two-step maximum likelihood method+method of moment

of $\widehat{ESD}_{MC}$ sensibly change moving from $\theta = 4$ to $\theta = 1/4$, holding fixed the other parameters; this is due to the different magnitude of the two values of θ here considered, whose choice was related to the different meaning of the copula parameter's values: for the Frank and Gauss copulas, changing the sign of θ means changing the sign to correlation while keeping its intensity fixed; this does not occur for the Plackett copula, whose parameter has to range within $\mathbb{R}^+$ (see Fig.2), leading to negative dependence if failing in (0, 1) and to positive dependence if larger than 1.

Although in this MC study, the results in terms of statistical performance of the inferential technique proposed in Sect. 5.3 are overall worse than those of the two more consolidated techniques described in Sects. 5.1 and 5.2, nevertheless the former can be still useful for providing starting values for the copula parameter to the maximization routines of the two latter two.

We also measured, on a selection of artificial settings, the total computational times in minutes required by each of the three estimation methods over the $S = 1000$ MC runs, which are displayed in Table 6. Although the magnitude of the computation time depends on the selected dependence structure, it can be noted that in each setting the two-step method of moment is by far the least time-consuming, followed

Table 6 Overall computation times in minutes for the three estimation methods for some artificial scenarios considered in the MC study

Type of dependence:	Gaussian			Frank			Plackett		
Sign of dependence	ML	TS	TSM	ML	TS	TSM	ML	TS	TSM
+	193.2	24.3	4.3	76.8	6.6	1.3	69.3	6.9	0.5
−	194.8	23.2	4.4	75.8	6.5	1.1	50.0	9.2	0.7

Here, $n = 100$; $m_1 = m_2 = 7$; $\xi_1 = 0.4, \pi_1 = 0.8, \xi_2 = 0.7, \pi_2 = 0.3$; "+" and "−" indicates the copula parameter value inducing positive and negative dependence, respectively (see Tables 5a, 5b, 5c). For TSM, $\epsilon = 10^{-7}$. Simulations run on RStudio Cloud

by the two-step maximum likelihood method, and the full maximum likelihood method. We remark that the suggested estimation procedure is faster when moving from the Gaussian copula to the other two dependence structures.

7 Empirical analysis

In this section, we consider an application of the inferential techniques of Sect. 5 to real data, specifically, the survey data coming from the 2000 International Social Survey Programme (ISSP), which addressed the topic of attitudes to environmental protection and preferred government measures for environmental protection. The `prefmod` package (Hatzinger and Dittirch 2012) comprises the raw data structured as a dataset with 1, 595 complete observations (one for each respondent) on 11 variables, namely five socio-demographical variables (gender, location of residence, age, country, and education) and six items (with a 5-point rating scale, i.e Likert type). Respondents from Austria and Great Britain were asked about their perception of environmental dangers; the questions concerned air pollution caused by cars (variable CAR), air pollution caused by industry (IND), pesticides and chemicals used in farming (FARM), pollution of country's rivers, lakes, and streams (WATER), a rise in the world's temperature (TEMP), and modifying the genes of certain crops (GENE). The answers were given on a 5-point rating scale, with response categories: (1) extremely dangerous, (2) very dangerous, (3) somewhat dangerous, (4) not very dangerous, and (5) not dangerous at all for the environment. We focus on respondents from Austria only and on WATER and GENE items. The joint and marginal empirical distributions of the two items are reported in Table 7. By considering the ratings as numerical values, we can treat this distribution as a sample from a bivariate discrete rv; its sample correlation is equal to 0.2988.

The nature of the data suggests using CUB as parametric family for the marginal distributions; the Gaussian, Frank and Plackett copulas are assumed as dependence structures for the joint distribution. Table 8 summarizes the estimation results obtained by applying the full maximum likelihood method and both the two-step methods. From Table 8, it is easy to note that within each dependence structure, the three estimation methods provide estimate values for the marginal parameters which are very close to each other; the estimates of the dependence parameter are slightly different. For example, under the Gaussian dependence structure, the values

Table 7 Joint distribution of the items WATER and GENE for Austrian respondents, ISSP2000 dataset

x, y	1	2	3	4	5	Total
1	115	42	19	5	3	184
2	105	149	35	17	4	310
3	43	78	60	20	7	208
4	17	26	21	7	1	72
5	1	2	2	2	1	8
Total	281	297	137	51	16	782

Table 8 Bivariate CUB models' estimates of the joint distribution of the items WATER and GENE for several dependence structures and according to the inferential methods of section 5

		π_1	ξ_1	π_2	ξ_2	θ	ℓ
Gauss	ML	0.98325	0.69147	0.87660	0.78214	0.36066	−2013.163
	TSML	0.98751	0.69090	0.88231	0.77991	0.35789	−2013.213
	TSM	0.98751	0.69090	0.88231	0.77991	0.34327	−2013.300
Frank	ML	0.98769	0.69218	0.88256	0.78140	2.37862	−2008.555
	TSML	0.98751	0.69090	0.88231	0.77991	2.37859	−2008.576
	TSM	0.98751	0.69090	0.88231	0.77991	2.23910	−2008.725
Plackett	ML	0.98730	0.69211	0.88221	0.78094	3.31883	−2006.317
	TSML	0.98751	0.69090	0.88231	0.77991	3.31882	−2006.329
	TSM	0.98751	0.69090	0.88231	0.77991	2.95842	−2006.790

ML= full maximum likelihood method; TS = two-step maximum likelihood method; TSM = two-step maximum likelihood method+method of moment

of the estimates of the dependence parameter ρ_{Ga} range between 0.343 and 0.361—by the way, this confirms a positive and moderate dependence between the two observed variables, as suggested by the value of the sample correlation. Differences in terms of maximized log-likelihood emerge across the three dependence structures here examined. Among the three models, the bivariate CUB with Plackett copula (and parameters set equal to the corresponding MLEs) is that providing the best fit ($\ell = -2006.317$).

Alternative ways of comparing the goodness of fit of the three models can be taken up; for example, one can consider Aitchison distance between the empirical and theoretical probabilities $\delta = \sum\sum(\log(\hat{p}_{ij}/p_{ij}) - \bar{L})^2$, where $\bar{L} = \sum\sum\log(\hat{p}_{ij}/p_{ij})/k$, with k being the total number of points of the support of the bivariate rv. The minimum value of the Aitchison distance is achieved by the model with the Plackett copula ($\Delta = 3.830$); the model with the Frank copula provides $\Delta = 4.089$; the Gauss copula returns $\Delta = 4.390$. Other possible distances or divergences between discrete distributions are mentioned in Fossaluza et al. (2018).

An evaluation in absolute terms of the goodness-of-fit of this bivariate model can be carried out by resorting to the chi-squared statistic defined as $\chi^2 = \sum\sum(n_{ij}^{obs} - n_{ij}^{theo})^2/n_{ij}^{theo}$, with n_{ij}^{obs} and n_{ij}^{theo} indicating the observed and the theoretical frequencies, respectively. In order to accomplish the requirement that each n_{ij}^{theo} has to be not smaller than 5, we can proceed and collapse the last two ordered categories (4 and 5) for each variable, thus obtaining the new observed joint distribution in Table 9. There, we also reported the theoretical joint frequencies corresponding to the "best" bivariate CUB model, for which the Chi-squared statistic above takes the value 21.68, with a p-value 0.0168 (the test statistic, under the null hypothesis that the data come from the bivariate CUB model, asymptotically follows a chi-squared distribution with $16 - 1 - 5 = 10$ degrees of freedom). This means that the goodness of fit of the model is hardly satisfactory; at the significance level 1%, we do not reject the null hypothesis. Looking at Table 9, discrepancies between

Table 9 Observed and theoretical (within brackets) joint frequencies of the items WATER and GENE under the bivariate CUB model with Plackett copula after collapsing the last two ordered categories

x, y	1	2	3	≥ 4	Total
1	115 *(102.3)*	42 *(56.4)*	19 *(15.1)*	8 *(5.3)*	184 *(179.1)*
2	105 *(117.5)*	149 *(135.4)*	35 *(47.3)*	21 *(17.0)*	310 *(317.2)*
3	43 *(44.1)*	78 *(89.0)*	60 *(54.6)*	27 *(24.7)*	208 *(212.3)*
≥ 4	18 *(11.1)*	28 *(25.5)*	23 *(22.6)*	11 *(14.0)*	80 *(73.3)*
Total	281 *(275.0)*	297 *(306.3)*	137 *(139.6)*	67 *(61.1)*	782 *(782)*

observed and theoretical joint (and also marginal) frequencies are visible to the unaided eye.

One can also consider not using a parametric model for the two margins and estimating them nonparametrically. In the opposite direction, other families of bivariate distributions may be tested for fit improvement; for example, cumulative link models could be used for the univariate margins (Agresti and Kateri 2019). Reasonably, introducing covariates (gender, education, age, location of residence) for the marginal (and dependence) parameters would likely increase the fit. However, we remark that the aim of this section was not so much to evaluate the fitting of bivariate (CUB) models to real data as to illustrate and compare different estimation techniques, one of which derived through a correlation-matching procedure. Besides, we remark that pooling cells, although being a viable alternative to obtain accurate p-values in some instances, should be preferably performed before the analysis is made in order to obtain a statistic with the appropriate asymptotic reference distribution; otherwise, it may distort the purpose of the analysis (Maydeu-Olivares and García-Forero 2010). Alternatively, one could resort to resampling methods (e.g. bootstrap), but unfortunately, existing evidence suggests that resampling methods do not yield accurate p-values for the χ^2 statistic (Tollenaar and Mooijart 2003).

8 Conclusions

In this work, we showed how to build a joint probability distribution with assigned discrete point-scale margins enjoying a target (feasible) value of correlation. We proposed a two-step copula-based approach: first, one selects a copula function and constructs a bivariate distribution preserving the assigned margins; then, one adjust the value of the copula parameter in order to achieve the target correlation: this leads to the elaboration of an iterative procedure whose accuracy can be set a priori, differently from other approaches in the literature that are based on some rearrangement of very huge samples drawn independently from the two margins.

The approach is designed to work with any one-parameter copula family, provided that it encompasses the entire range of dependence, thus allowing the use of other dependence structures than the Gaussian, whose use in the social sciences has been overriding.

Although in this paper we considered three exchangeable and radially symmetric copulas, this feature is not necessary for the algorithm functioning. Moreover, being comprehensive is a property that makes the algorithm applicable to any feasible correlation value; however, if one is only concerned, for example, with positive correlations, then he/she can use non-comprehensive copulas as well, such as Gumbel or Clayton.

As said, the algorithm is specifically conceived for one-parameter copulas, but it can be extended to two or more-parameter copulas: in this case, some higher comoments need to be assigned along with linear correlation in order to calibrate the additional parameters.

The extension of the proposed procedure to dimension $d > 2$, discussed in Sect. 3.4, is not straightforward, being limited by theoretical rather than computational reasons, related to the properties of Pearson's correlation matrix for a d-variate random vector.

Future research will explore these aspects.

9 Supplementary material

The `R` code implementing the algorithm of Section 3, along with the example of Section 4, is provided as supplementary material here: https://tinyurl.com/ASTB-D-19-00215.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Agresti, A., Kateri, M.: The class of CUB models: statistical foundations, inferential issues and empirical evidence. Statist. Meth. Appl. **28**, 445–449 (2019)

Agresti, A.: Analysis of Ordinal Categorical Data. Wiley, Hoboken (2010)

Andreis, F., Ferrari, P.A.: On a copula model with CUB margins. Quad. Stat. **15**, 33–51 (2013)

Barbiero, A., Ferrari, P.A.: GenOrd: Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions. R package version 1.4.0 (2015)

Barbiero, A., Ferrari, P.A.: An R package for the simulation of correlated discrete variables. Comm. Stat. Simul. Comput. **46**(7), 5123–5140 (2017)

Barbiero, A.: Inducing a desired value of correlation between two point-scale variables, in: ASMOD 2018 Proceedings of the International Conference on Advances in Statistical Modelling of Ordinal Data, fedOA Press, Naples, pp. 45-52 (2018)
Bergsma, W., Rudas, T. (2002). Conditional and marginal association in contingency tables. (2018). URL=http://yaroslavvb.com/papers/bergsma-conditional.pdf
Carifio, J., Perla, R.: Resolving the 50-year debate around using and misusing Likert scales. Med. Educ. **42**, 1150–1152 (2008)
Cario, M., Nelson, B.: Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, Tech. rep (1997)
Chaganty, N.R., Joe, H.: Range of correlation matrices for dependent Bernoulli random variables. Biometrika **93**, 197–206 (2006)
Corduas, M.: Analyzing bivariate ordinal data with CUB margins. Stat. Model. **15**(5), 411–432 (2011)
Demirtas, H.: A method for multivariate ordinal data generation given marginal distributions and correlations. J. Stat. Comput. Simul. **76**(11), 1017–1025 (2006)
Demirtas, H., Yavuz, Y.: Concurrent generation of ordinal and normal data. J. Biopharm. Stat. **25**(4), 635–650 (2015)
Demirtas, H., Vardar-Acar, C.: Anatomy of correlational magnitude transformations in latency and discretization contexts in Monte-Carlo studies. In: Chen, D.G., Chen, J.D. (eds.) Monte-Carlo Simulation-Based Statistical Modeling, pp. 59–84. Springer, Singapore (2017)
Demirtas, H.: Inducing any feasible level of correlation to bivariate data with any marginals. Amer. Stat. **73**(3), 273–277 (2019)
Devroye, L.: Non-Uniform Random Variate Generation. Springer, New York (1986)
Faust, K., Wasserman, S.: Correlation and association models for studying measurements on ordinal relations. Sociol. Methodol. **23**, 177–215 (1993)
Ferrari, P.A., Manzi, G.: Nonlinear principal component analysis as a tool for the evaluation of customer satisfaction. Qual. Technol. Quant. Manag. **7**(2), 117–132 (2010)
Ferrari, P.A., Barbiero, A.: Simulating ordinal data. Multivar. Behav. Res. **47**(4), 566–589 (2012)
Foldnes, N., Olsson, U.H.: A simple simulation technique for nonnormal data with prespecified skewness, kurtosis, and covariance matrix. Multivariate Behav. Res. **51**(2–3), 207–219 (2016)
Fossaluza, V., Esteves, L.G., Pereira, CAd.B.: Estimating Multivariate Discrete Distributions Using Bernstein Copulas. Entropy **20**(3), 194 (2018)
Hatzinger, R., Dittrich, R.: prefmod: an R package for modeling preferences based on paired comparisons, rankings, or ratings. J. Stat. Softw. **48**(10), 1–31 (2012)
Oh, D.H., Patton, A.J.: Modeling dependence in high dimensions with factor copulas. J. Bus. Econom. Stat. **35**(1), 139–154 (2017)
Iannario, M., Piccolo, D.: A new statistical model for the analysis of customer satisfaction. Qual. Technol. Quant. Manag. **7**(2), 149–168 (2010)
Iannario, M., Piccolo, D.: CUB models: Statistical methods and empirical evidence, in: Kenett R and Salini S Eds., Modern Analysis of Customer Surveys: with applications using R, 231-258 (2012)
Kruskal, W.H., Goodman, L.: Measures of association for cross classifications. J. Am. Stat. Assoc. **49**(268), 732–764 (1954)
Joe, H., Xu, J.J.: The estimation method of inference functions for margins for multivariate models, p. 166. Technical Report, UBC, Department of Statistics (1996)
Joe, H.: Dependence Modeling with Copulas. CRC Press, Boca Raton, FL (2014)
Lancaster, H.O.: Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika **44**, 289–292 (1957)
Lee, A.J.: Some methods for generating correlated categorical variates. Comput. Stat. Data Anal. **26**, 133–148 (1997)
Lee, Y., Kaplan, D.: Generating multivariate ordinal data via entropy principles. Psychometrika **83**(1), 156–181 (2018)
Leonov, S., Qaqish, B.: Correlated endpoints: simulation, modeling, and extreme correlations. Stat. Papers **61**(2), 741–766 (2020)
Liu, I., Agresti, A.: The analysis of ordered categorical data: an overview and a survey of recent developments. Test **14**, 1–73 (2005)
Madsen, L., Dalthorp, D.: Simulating correlated count data. Environ. Ecol. Stat. **14**(2), 129–148 (2007)
Maydeu-Olivares, A., García-Forero, C.: Goodness-of-fit testing. Int. Encycl. Educ. **2010**(7), 190–196 (2010)

McNeil, A., Frey, R., Embrechts, P.: Quantitative risk management. Princeton Series in Finance, Princeton, Concepts, Techniques and Tools (2005)
Mittelhammer, R.C.: Mathematical Statistics for Economics and Business, 2nd edn. Springer, New York (2013)
Nelsen, R.B.: Discrete bivariate distributions with given marginals and correlation. Comm. Stat. Simul. Comput. **16**(1), 199–208 (1987)
Nelsen, R.B.: An Introduction to Copulas. Springer Verlag, New York (2006)
Nešlehová, J.: On rank correlation measures for non-continuous random variables. J. Multivar. Anal. **98**(3), 544–567 (2007)
Norman, G.: Likert scales, levels of measurement and the laws of statistics. Adv. Health Sci. Educ. **15**(5), 625–632 (2010)
Okhrin, O., Tetereva, A.: The realized hierarchical Archimedean Copula in risk modelling. Econometrics **5**(2), 26 (2017)
Patton, A.L.: Estimation of multivariate models for time series of possibly different lengths. J. Appl. Econometr. **21**(2), 147–173 (2006)
Piccolo, D.: On the moments of a mixture of uniform and shifted binomial random variables. Quad. Stat. **5**, 85–104 (2003)
Ruiz, M.D., Hüllermeier, E.: A formal and empirical analysis of the fuzzy gamma rank correlation coefficient. Inform. Sci. **206**, 1–17 (2012)
Salvemini, T.: Sugli indici di omofilia. Supplemento statistico Nuovi Problemi **5**, 105–115 (1939)
Scarsini, M., Shaked, M.: Positive dependence orders: A survey. In Athens conference on applied probability and time series analysis (pp. 70-91). Springer, New York, NY (1996)
Tollenaar, N., Mooijart, A.: Type I errors and power of the parametric goodness-of-fit test: full and limited information. British J. Math. Statist. Psych. **56**(2), 271–288 (2003)
Xiao, Q.: Generating correlated random vector involving discrete variables. Comm. Stat. Theory Methods **46**(4), 1594–1605 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.