Austrian Journal of Statistics
April 2020, Volume 49, 9–18.
http://www.ajs.or.at/
doi:10.17713/ajs.v49i4.1116

Inducing a Target Association between Ordinal Variables by Using a Parametric Copula Family

Alessandro Barbiero
Università degli Studi di Milano

Abstract

The need for building and generating statistically dependent random variables arises in various fields of study where simulation has proven to be a useful tool. In this work, we present an approach for constructing ordinal variables with arbitrarily assigned marginal distributions and value of association or correlation, expressed in terms of either Goodman and Kruskal's gamma or Pearson's linear correlation. The approach first constructs a class of bivariate copula-based distributions matching the assigned margins, and then, within this class, identifies the distribution matching the assigned association or correlation, by calibrating the copula parameter. A numerical example and a possible application are illustrated.

Keywords: bivariate normal distribution, discretization, gamma coefficient, latent variable, ordinal association.

1. Introduction

The need for building and drawing samples from statistically dependent random variables emerges in various fields of study where simulation has proven to be a powerful tool. The ability to simulate data resembling the observed data is fundamental to compare and investigate the behaviour of statistical procedures when analytical results are not derivable or are cumbersome to derive.

Many datasets, especially those arising in the social sciences, often contain ordinal variables. Sometimes they are genuine ordered assessments (judgements, preferences, degree of liking, etc.) whereas in other circumstances they are discretized or categorized for convenience (e.g., age of people in classes or education achievement). There are several statistical models and techniques that can be employed for handling multivariate ordinal data without trying to quantify their ordered categories: Agresti (2010) gives 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 (Faust and Wasserman 1993). Alternatively, one can code the ordered categories as integers numbers $(1, 2 \ldots, m)$: this amounts to assuming that the categories are evenly spaced.

In this work, we present an approach for constructing ordinal variables with arbitrary marginal distributions and assigned value of association, expressed in terms of either Goodman and Kruskal's gamma or Pearson's linear correlation. Proposals that aim at solving similar problems have been already suggested by Lee (1997), when dealing with ordinal variables and Goodman and Kruskal's gamma, and by Demirtas (2006); Madsen and Dalthorp (2007); Ferrari and Barbiero (2012) for ordinal (and count) variables and Pearson's correlation.

The rest of the paper is structured as follows. Section 2 states the statistical problem. Section 3 proposes a solution employing bivariate copula functions. Section 4 presents a numerical example. Section 5 illustrates an application involving real data. Section 6 provides final remarks.

2. Statement of the problem

Following Barbiero (2019), we consider two ordinal random variables (rvs), X and Y, with h and k ordered categories, respectively, with marginal distributions $p_{i\cdot} = P(X = x_i), i = 1, \dots, h$, and $p_{\cdot j} = P(Y = y_j), j = 1, \dots, k$. We want to determine *some* joint probability distribution $p_{ij} = P(X = x_i, Y = y_j)$, $i = 1, \dots, h$, $j = 1, \dots, k$, such that its margins are actually $p_{i\cdot}$ and $p_{\cdot j}$, and with an assigned level of association.

Being X and Y ordinal variables, the association can be naturally expressed through the Goodman and Kruskal's gamma coefficient (Goodman and Kruskal 1954). Considering two independent realizations (X_s, Y_s) and (X_t, Y_t) of (X, Y), Goodman and Kruskal's gamma is defined as

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}, \tag{1}$$

where Π_c is the probability of concordance:

$$\Pi_c = Pr\left\{X_s < X_t \text{ and } Y_s < Y_t\right\} + Pr\left\{X_s > X_t \text{ and } Y_s > Y_t\right\},$$

and Π_d the probability of discordance:

$$\Pi_d = Pr\left\{X_s < X_t \text{ and } Y_s > Y_t\right\} + Pr\left\{X_s > X_t \text{ and } Y_s < Y_t\right\},$$

and Π_c and Π_d can be conveniently expressed in terms of the joint probabilities p_{ij}:

$$\Pi_c = \sum_{r=1}^{h}\sum_{c=1}^{k} p_{rc}\left(\sum_{i<r}\sum_{j<c} p_{ij} + \sum_{i>r}\sum_{j>c} p_{ij}\right), \quad \Pi_d = \sum_{r=1}^{h}\sum_{c=1}^{k} p_{rc}\left(\sum_{i<r}\sum_{j>c} p_{ij} + \sum_{i>r}\sum_{j<c} p_{ij}\right).$$

γ take values in the $[-1, +1]$ interval; in particular, the values -1, 0, and $+1$ are attained when $\Pi_c = 0$, $\Pi_c = \Pi_d$, $\Pi_d = 0$, respectively. However, a value of γ equal to ± 1 implies that the relationship between X and Y is monotone, but not strictly monotone. Moreover, γ generally takes on larger absolute values than other association measures: to overcome this shortcoming, a modification of Goodman and Kruskal's coefficient has been proposed in Kvålseth (2017). We notice that if the rvs X and Y are continuous, then the gamma coefficient defined in (1) boils down to Kendall's rank correlation τ (Kendall 1938).

If we treat X and Y as point-scale discrete variables, by assigning the first h and k positive integers, respectively, to their ordered categories, then we can use Pearson's correlation coefficient as a measure of association:

$$\rho = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \tag{2}$$

with $\mathbb{E}(X) = \sum_{i=1}^{h} i p_{i\cdot}$, $\text{Var}(X) = \sum_{i=1}^{h}(i - \mathbb{E}(X))^2 p_{i\cdot}$ (analogous definitions hold for Y), and $\mathbb{E}(XY) = \sum_{i=1}^{h}\sum_{j=1}^{k} ij p_{ij}$. Like γ, also Pearson's correlation takes values in the $[-1, +1]$ interval; however, given two marginal distributions and a value $\rho \in [-1, +1]$, it is not always

possible to construct a joint distribution with those assigned margins, whose correlation is equal to the assigned ρ (McNeil, Frey, and Embrechts 2005). In more detail, the attainable correlations form a closed interval $[\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} < 0 < \rho_{\max}$. The minimum correlation $\rho_{\min}$ is attained if and only if X and Y are countermonotonic; the maximum correlation $\rho_{\max}$ is attained if and only if X and Y are comonotonic. Moreover, $\rho_{\min} = -1$ if and only if X and $-Y$ are of the same type, and $\rho_{\max} = 1$ if and only if X and Y are of the same type. A correlation value ρ is said "feasible" if it falls within $[\rho_{\min}, \rho_{\max}]$.

3. A two-step solution employing a parametric copula family

Finding a joint probability distribution with assigned margins and a desired (feasible) value of association is mathematically equivalent to solving a system in $h \times k$ unknowns, the p_{ij}, belonging to the standard simplex, subject to $h + k - 1$ constraints corresponding to the assigned margins and one further constraint dictated by the desired association. This system, when the number of categories h or k is greater than 2, has infinite solutions, which can be recovered more easily when using Pearson's correlation as a measure of association, being it a linear function in the p_{ij} (the p_{ij} appear – with power 1 – only in the term $\mathbb{E}(XY)$ of Equation 2).

Here we propose an approach to identify just one solution, i.e., one joint distribution from among all the distributions satisfying the requirements on the margins and association value. This procedure relies on one-parameter bivariate copulas, which allow us to split the original problem into two sequential steps: first, identifying a class of joint distributions respecting the assigned margins; then, within this class, finding the joint distribution matching the desired level of association by properly calibrating the copula parameter.

3.1. Selecting a class of joint distributions having the pre-specified margins

As for the first step, if F_1 and F_2 are the cumulative distribution functions of the two rvs X and Y, $F_1(x) = Pr\{X \leq x\}$ and $F_2(y) = Pr\{Y \leq y\}$, and $C(u, v; \theta)$ is a bivariate parametric copula family, characterized by some scalar parameter θ, the function

$$F(x, y) = C(F_1(x), F_2(y); \theta), \quad x, y \in \mathbb{R}, \tag{3}$$

defines a valid joint cumulative distribution function, whose margins are exactly F_1 and F_2 (Sklar 1959). This result keeps holding if X and Y are ordinal or discrete; in this case, the marginal cumulative probabilities are $F_{i\cdot} = Pr\{X \leq x_i\}$ and $F_{\cdot j} = Pr\{Y \leq y_j\}$, the joint cumulative probabilities can be computed from the analog of Equation (3), $F_{i,j} = C(F_{i\cdot}, F_{\cdot j}; \theta)$, and the joint probabilities are derived as:

$$p_{ij} = F_{i,j} - F_{i-1,j} - F_{i,j-1} + F_{i-1,j-1}, \tag{4}$$

for $i = 1, \ldots, h; j = 1, \ldots, k$.

In order to induce any feasible value of association between the two discrete margins, we have further to impose that the copula $C(u, v; \theta)$ is "comprehensive", i.e., by varying θ, it encompasses the entire range of dependence, from perfect negative dependence to perfect positive dependence passing through independence. The Gauss, Frank and Plackett copulas are well-known examples of comprehensive copulas: Table 1 displays for each of them the expression of the copula function and the value of their scalar parameter leading to the countermonotonicity, independence and comonotonicity copula as special cases.

3.2. Inducing the desired value of association

As for the second step, the association between X and Y now depends only on the copula parameter θ; this relationship may be written in an analytical or, more frequently, numerical

Table 1: Three parametric comprehensive bivariate copulas: expressions of the copula and parameter values for which the copula reduces to the countermonotonicity, independence, and comonotonicity copula.

copula	function $C(u_1, u_2)$	counterm.	indep.	comon.
Gauss	$\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho_{Ga}^2}} e^{-\frac{s_1^2 - 2\rho_{Ga} s_1 s_2 + s_2^2}{2(1-\rho_{Ga}^2)}} \mathrm{d}s_1 \mathrm{d}s_2$	$\rho_{Ga} = -1$	$\rho_{Ga} = 0$	$\rho_{Ga} = +1$
Frank	$-\frac{1}{\kappa} \ln \left[1 + \frac{(e^{-\kappa u_1} - 1)(e^{-\kappa u_2} - 1)}{e^{-\kappa} - 1} \right]$	$\kappa \to -\infty$	$\kappa \to 0$	$\kappa \to +\infty$
Plackett	$\frac{1+(\theta-1)(u_1+u_2) - \sqrt{[1+(\theta-1)(u_1+u_2)]^2 - 4\theta(\theta-1)u_1 u_2}}{2(\theta-1)}$	$\theta \to 0$	$\theta \to 1$	$\theta \to +\infty$

form, say $\gamma = f(\theta)$, or $\rho = g(\theta)$. Since the function f (or g) is not usually analytically invertible, inducing a desired feasible value of association, 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 $f(\theta) - \gamma = 0$ (or $g(\theta) - \rho = 0$). If γ (or ρ) is a monotone increasing function of the copula parameter, and this is often the case (e.g., for the Gauss, Frank, and Plackett copulas), one can implement some iterative root-finding procedure that is more efficient than the standard bisection method. For discrete random variables, several proposals have been suggested for matching a desired value of Pearson's correlation when the copula is Gaussian, see Demirtas (2006); Madsen and Dalthorp (2007); Ferrari and Barbiero (2012). An `R` implementation based on Ferrari and Barbiero (2012) is presented in Barbiero and Ferrari (2017).

Basically, one can start by setting a trial value of the copula parameter θ and then compute the corresponding cumulative distribution function (3), probability mass function (4), and association (or correlation) value for the corresponding bivariate ordinal distribution – equations (1) or (2). If the resulting value of association (correlation) is equal to the assigned value apart from an arbitrary small absolute difference ϵ, the algorithm stops, otherwise, one has to iteratively update the value of θ (for example, simply using some linear interpolation) till the corresponding value of association (or correlation) converges to the target one. If the selected copula is the Gaussian one, with parameter ρ_{Ga}, then a convenient choice of its initial trial value can be the target ρ itself, in case the matching is on the correlation coefficient, or the value $\rho_{Ga} = \sin(\pi\gamma/2)$, in case the matching is on γ (we recall that for the Gaussian copula, the following relationship holds between rank and linear correlation: $\tau_{Ga} = \frac{2}{\pi} \arcsin \rho_{Ga}$).

Then, simulating from the selected joint distribution is straightforward, by resorting to preliminary simulation of copulas or more easily to a direct inversion algorithm (Devroye 1986; Lee 1997).

4. A numerical example

Let us consider two ordinal variables X and Y whose assigned marginal distributions are as follows:

$$P(X = x_1) = 1/15, P(X = x_2) = 2/15, P(X = x_3) = 1/5, P(X = x_4) = 4/15, P(X = x_5) = 1/3;$$

$$P(Y = y_1) = 1/4, P(Y = y_2) = 1/2, P(Y = y_3) = 1/4.$$

Based on the two margins, one can build the cograduation and countergraduation tables, which are reported in Table 2. Then, assuming that the ordered categories of X and Y are evenly spaced and then can be substituted by the first $h = 5$ and $k = 3$ positive integers,

one can compute the minimum and maximum attainable correlations, which are equal to $\rho_{\min} = -0.8693183$ and $\rho_{\max} = 0.8693183$. We notice that $\rho_{\min}$ is different from -1 and $\rho_{\max}$ is different from $+1$; the fact that $\rho_{\min} = -\rho_{\max}$ is due to the symmetry of one of the two marginal distributions, in this case that of Y.

Let us now assume that we want to construct a bivariate distribution with the above margins and a feasible value of correlation $\rho = 0.5$, by employing the Gauss copula, which is characterized by the scalar parameter $\rho_{Ga} \in [-1, +1]$. Then one applies the algorithm sketched in the previous section in order to find the corresponding value of ρ_{Ga}. A trial value of ρ_{Ga} can be the target value $\rho = 0.5$ itself. Setting a maximum tolerated absolute error $\epsilon = 10^{-7}$, after 4 iterations the final value of ρ_{Ga} is computed as 0.5891106 and the resulting bivariate distribution is displayed in Table 3. If we move to the Frank copula, it is not immediate to set a trial value for κ; however, since we know that positive values of κ lead to positive correlation, one should use a positive number as a starting value, say $\kappa = 1$; the final value of the parameter ensuring the desired value of correlation is $\kappa = 4.178493$ (6 iterations). The resulting bivariate distribution is displayed in Table 4. Finally, we consider the Plackett copula: since values greater than 1 of its parameter θ are responsible for positive correlation, one should select the starting value of θ in $(1, +\infty)$, say $\theta = 2$; the final value of θ provided by the iterative search procedure after 8 iterations is 6.877371. The resulting bivariate distribution is displayed in Table 5. Notice the differences among homologous joint probabilities across Tables 3, 4, and 5.

Table 2: Cograduation (left) and countergraduation (right) tables for the marginal distributions of the numerical example.

X, Y	y_1	y_2	y_3	tot
x_1	1/15	0	0	1/15
x_2	2/15	0	0	2/15
x_3	1/20	3/20	0	1/5
x_4	0	4/15	0	4/15
x_5	0	1/12	1/4	1/3
tot	1/4	1/2	1/4	1

X, Y	y_1	y_2	y_3	tot
x_1	0	0	1/15	1/15
x_2	0	0	2/15	2/15
x_3	0	3/20	1/20	1/5
x_4	0	4/15	0	4/15
x_5	1/4	1/12	0	1/3
tot	1/4	1/2	1/4	1

Table 3: Joint distribution obtained by combining the two margins of Table 3 with a Gaussian copula, with correlation $\rho = 0.5$.

X, Y	y_1	y_2	y_3	tot
x_1	0.0474	0.0183	0.0010	1/15
x_2	0.0661	0.0606	0.0067	2/15
x_3	0.0657	0.1118	0.0225	1/5
x_4	0.0500	0.1575	0.0593	4/15
x_5	0.0209	0.1519	0.1606	1/3
tot	1/4	1/2	1/4	1

Figure 1 displays the graph of the function $\rho = g(\theta)$, linking the correlation coefficient to the copula parameter, for the three classes of copula-based bivariate distributions analyzed in this example. Although the function is monotone increasing in all the three cases, one can notice the different shapes of the curve moving from the top to the bottom graph (almost linear, "S"-shaped, concave), which heavily depends on the range of the copula parameter (limited, unlimited to both sides, unlimited to the right).

Alternatively, preserving the original ordinal nature of the two variables, one can assign a target value to the gamma coefficient, by considering the usual $[-1, +1]$ interval, say $\gamma = 0.5$, and then recover the value of the copula parameter ensuring this level of association given the choice of margins. Focusing for example on the Gauss copula, by using the root-search

Table 4: Joint distribution obtained by combining the two margins of Table 3 with a Frank copula, with correlation $\rho = 0.5$.

X, Y	y_1	y_2	y_3	tot
x_1	0.0417	0.0227	0.0022	1/15
x_2	0.0699	0.0568	0.0066	2/15
x_3	0.0713	0.1095	0.0191	1/5
x_4	0.0471	0.1609	0.0587	4/15
x_5	0.0199	0.1500	0.1634	1/3
tot	1/4	1/2	1/4	1

Table 5: Joint distribution obtained by combining the two margins of Table 3 with a Plackett copula, with correlation $\rho = 0.5$.

X, Y	y_1	y_2	y_3	tot
x_1	0.0440	0.0193	0.0033	1/15
x_2	0.0720	0.0530	0.0083	2/15
x_3	0.0679	0.1129	0.0193	1/5
x_4	0.0428	0.1707	0.0531	4/15
x_5	0.0232	0.1441	0.1660	1/3
tot	1/4	1/2	1/4	1

algorithm of Section 3.2 and setting the initial value of ρ_{Ga} to $\sin(\pi \cdot 0.5/2) = 0.7071068$, we obtain that the value $\rho_{Ga} = 0.4758166$ allows us to recover a bivariate ordinal distribution with the desired features.

5. An application to real data

In a now classic study of mental health in Manhattan, New York, Srole and Fischer (1978) explore the relationship, among others, between mental impairment (Y) and parents' socioeconomic status (X), both measured on an ordinal scale. Table 6, from that study, has been used extensively to illustrate the utility and application of models for ordered categorical data.

Table 6: The Midtown Manhattan Study: Mental Health and Parents' Socioeconomic Status (Srole and Fischer, 1978). In normal font, the observed joint frequencies; in italic, between brackets, the expected joint frequencies under the bivariate ordinal model obtained by matching the empirical marginal distributions and the sample Goodman and Kruskal's gamma coefficient.

	Mental Health				
Parents' Socioeconomic Status	Well	Mild symptom formation	Moderate symptom formation	Impaired	total
A (high)	64 (*67.62*)	94 (*102.02*)	58 (*50.19*)	46 (*42.17*)	262
B	57 (*53.05*)	94 (*93.30*)	54 (*50.90*)	40 (*47.75*)	245
C	57 (*55.74*)	105 (*106.91*)	65 (*62.00*)	60 (*62.35*)	287
D	72 (*65.70*)	141 (*138.60*)	77 (*86.04*)	94 (*93.66*)	384
E	36 (*39.11*)	97 (*91.64*)	54 (*61.30*)	78 (*72.96*)	265
F (low)	21 (*25.78*)	71 (*69.53*)	54 (*51.58*)	71 (*70.11*)	217
total	307	602	362	389	1660

We consider this dataset and compute the empirical marginal distributions for the variables

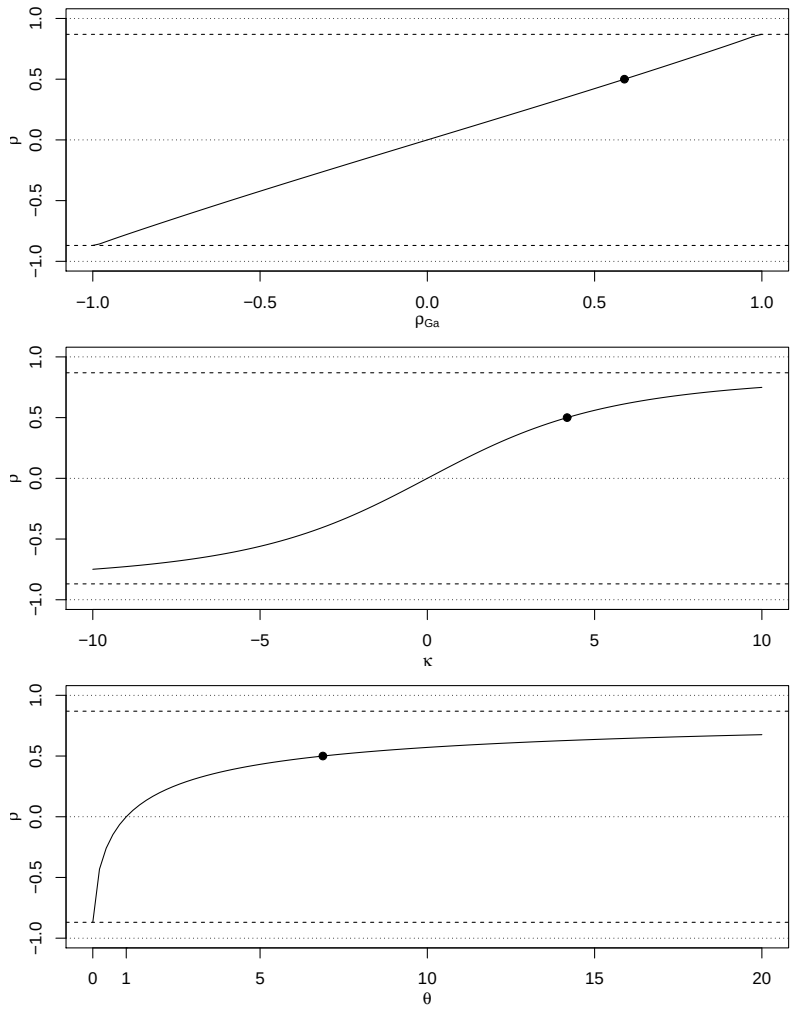

Figure 1: Relationship between the linear correlation ρ and the copula parameter θ for three different classes of bivariate copula-based ordinal distributions, with marginal distributions specified in Section 4. From the top to the bottom graph, the copulas are Gauss, Frank and Plackett. In each graph, the horizontal dashed lines represent the actual bounds for the correlation coefficient; the dotted lines indicate the "standard" -1 and $+1$ bounds; the thickened points are drawn in correspondence to $\rho = 0.5$.

X and Y and the sample Goodman and Kruskal's gamma coefficient, equal to 0.15429. Then, we construct a bivariate ordinal rv matching these two margins and their ordinal association value, following the lines of Section 3, by choosing a Gaussian dependence structure. The corresponding value of ρ_{Ga} is computed as 0.16674 (4 iterations required with a maximum tolerated absolute error equal to 10^{-7}). Comparing the observed and expected joint frequencies under this bivariate ordinal distribution (see Table 6) leads us to believe that this latter fits the data more than adequately. Then, we carry out the following Monte Carlo simulation plan: we simulate a huge number $S = 10,000$ of samples of the same size $n = 1,660$ of the original dataset from the selected rv: this way, we are producing "replicates" of the original sample, in the sense that they preserve some of its main features. This simulation plan can be regarded to as a sort of resampling technique applied to a bivariate sample: we draw simple random samples from a bivariate rv whose margins are exactly the empirical margins of the sample, whose ordinal association is the sample association computed on the sample, and whose dependence structure is Gaussian.

The results of the simulation plan are summarized in Table 7, displaying the sample mean and standard deviation of each joint frequency across the $10,000$ simulated datasets. Obviously,

due to the law of large numbers, the Monte Carlo averages of the joint frequencies are very close to the corresponding expected values displayed in Table 6 between brackets. The Monte Carlo mean of Goodman and Kruskal's gamma is about 0.15432 with a standard deviation of 0.02545. We can notice that this latter Monte Carlo quantity is very close to the asymptotic standard error 0.02482 which can be computed through the formula of the asymptotic variance of the sample gamma coefficient of Equation (2.7) contained in Goodman and Kruskal (1972). Such a Monte Carlo experiment can be also used for detecting the presence of outlying cells in the dataset at study. For each cell (i, j), one can compute the standardized or the adjusted residual (Haberman 1973), by considering on the observed and theoretical joint frequencies (the figures in normal and italic font in Table 6, respectively). Based on the $10,000$ simulated datasets, one can then build a sort of bootstrapped distribution for each adjusted residual, instead of just computing a single value, and compare it to its asymptotic distribution, i.e., the standard normal.

Analogous simulation studies can be carried out by selecting a different dependence structure (for example, by employing the Frank and Plackett copulas).

Table 7: Simulation study: Monte Carlo average and standard deviation of the joint frequencies of $10,000$ replicates of the Midtown Manhattan Study dataset of Table 6, with the same sample size $n = 1,660$, drawn from a bivariate distribution preserving its marginal distributions and ordinal association.

	Y				
X	y_1	y_2	y_3	y_4	total
x_1	67.5 (8.0)	102.1 (9.8)	50.2 (7.0)	42.1 (6.4)	261.9 (14.7)
x_2	53.1 (7.1)	93.2 (9.4)	51.0 (7.1)	47.7 (6.8)	245.0 (14.5)
x_3	55.6 (7.4)	106.9 (10.1)	62.0 (7.7)	62.4 (7.7)	287.0 (15.3)
x_4	65.8 (8.0)	138.6 (11.3)	86.0 (9.0)	93.7 (9.3)	384.0 (17.2)
x_5	39.2 (6.2)	91.5 (9.3)	61.3 (7.8)	73.0 (8.3)	265.0 (25.0)
x_6	25.8 (5.0)	69.6 (8.1)	51.5 (7.1)	70.3 (8.2)	217.2 (13.7)
total	307.0 (15.7)	601.9 (19.6)	362.0 (16.9)	389.1 (17.2)	1660

6. Conclusions

We described a procedure for constructing a bivariate ordinal random variable matching two assigned marginal distributions and a feasible assigned value of association or correlation, expressed in terms of Goodman and Kruskal's gamma coefficient or Pearson's correlation. The procedure relies on a parametric copula function (used for matching the margins) and a root-searching algorithm (used for matching the pairwise association or correlation by varying the copula parameter). A numerical example and a possible application are illustrated. The procedure has been implemented in the `R` environment (R Core Team 2019) and relevant code will be made freely available.

We remark that the choice of the parametric copula to employ should be based on goodness-of-fit arguments if one is interested in generating replicates of a given dataset; otherwise, if one just needs to draw samples with assigned margins and value of a bivariate association measure, then the Gaussian copula is probably the most convenient choice, which usually requires the smallest number of iterations for the root-searching algorithm. However, we remind that the three copulas considered in this paper possess similar features: in dimension two, they are all comprehensive, exchangeable, radially symmetric, and tail-independent.

Finally, we are aware that the concept of copula is not so natural for discrete/ordinal variables and that it may raise serious concerns in terms of model identifiability and estimation (Faugeras 2017); nevertheless, it can be still effectively used as a valuable tool for building and simulating a model with prescribed features.

References

Agresti A (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons, New York.

Barbiero A (2019). "Inducing a Target Association between Ordinal Variables by Using a Parametric Copula Family." In *Computer Data Analysis and Modeling: Stochastics and Data Science*, pp. 13–16. Belarusian State University.

Barbiero A, Ferrari PA (2017). "An R Package for the Simulation of Correlated Discrete Variables." *Communications in Statistics-Simulation and Computation*, **46**(7), 5123–5140.

Demirtas H (2006). "A Method for Multivariate Ordinal Data Generation Given Marginal Distributions and Correlations." *Journal of Statistical Computation and Simulation*, **76**(11), 1017–1025.

Devroye L (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

Faugeras OP (2017). "Inference for Copula Modeling of Discrete Data: A Cautionary Tale and some Facts." *Dependence Modeling*, **5**(1), 121–132.

Faust K, Wasserman S (1993). "Correlation and Association Models for Studying Measurements on Ordinal Relations." *Sociological Methodology*, **23**, 177–215.

Ferrari PA, Barbiero A (2012). "Simulating Ordinal Data." *Multivariate Behavioral Research*, **47**(4), 566–589.

Goodman LA, Kruskal WH (1954). "Measures of Association for Cross Slassifications." *Journal of the American Statistical Association*, **49**, 732–764.

Goodman LA, Kruskal WH (1972). "Measures of Association for Cross Classifications, IV: Simplification of Asymptotic Variances." *Journal of the American Statistical Association*, **67**(338), 415–421.

Haberman SJ (1973). "The Analysis of Residuals in Cross-classified Tables." *Biometrics*, **29**(1), 205–220.

Kendall MG (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**(1/2), 81–93.

Kvålseth TO (2017). "An Alternative Measure of Ordinal Association as a Value-validity Correction of the Goodman–Kruskal Gamma." *Communications in Statistics-Theory and Methods*, **46**(21), 10582–10593.

Lee AJ (1997). "Some Methods for Generating Correlated Categorical Variates." *Computational Statistics and Data Analysis*, **26**(2), 133–148.

Madsen L, Dalthorp D (2007). "Simulating Correlated Count Data." *Environmental and Ecological Statistics*, **14**(2), 129–148.

McNeil A, Frey R, Embrechts P (2005). *Quantitative Risk Management. Concepts, Techniques and Tools*. Princeton Series in Finance, Princeton.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Sklar M (1959). "Fonctions de Répartition à n Dimensions et Leurs Marges." *Publications de l'Institut de statistique de l'Université de Paris*, **8**, 229–231.

Srole LE, Fischer AK (1978). *Mental Health in the Metropolis: The Midtown Manhattan Study*. Rev. edition. New York U Press.

Affiliation:

Alessandro Barbiero
Department of Economics, Management and Quantitative Methods
Università degli Studi di Milano
via Conservatorio 7, 20122 Milan, Italy
E-mail: alessandro.barbiero@unimi.it
URL: https://www.unimi.it/en/ugov/person/alessandro-barbiero

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 49
April 2020

http://www.ajs.or.at/
http://www.osg.or.at/

Submitted: 2019-12-15
Accepted: 2020-03-03