Scientia Iranica E (2018) 25(1), 386–397

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

A proposal for modeling and simulating correlated discrete Weibull variables

A. Barbiero*

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

Received 20 December 2015; received in revised form 10 October 2016; accepted 21 November 2016

KEYWORDS
Correlated counts;
Gaussian copula;
Parameter estimation;
Stochastic simulation.

Abstract. Researchers in the applied sciences are often concerned with multivariate random variables. In particular, multivariate discrete data often arise in many fields (statistical quality control, biostatistics, failure and reliability analysis, etc.) and modeling such data is a relevant task, as well as simulating correlated discrete data satisfying some specific constraints. Here, we consider the discrete Weibull distribution as an alternative to the popular Poisson random variable and propose a procedure for simulating correlated discrete Weibull random variables with marginal distributions and correlation matrix assigned by the user. The procedure relies upon the Gaussian copula model and an iterative algorithm for recovering the proper correlation matrix for the copula ensuring the desired correlation matrix on the discrete margins. A simulation study is presented, which empirically assesses the performance of the procedure in terms of accuracy and computational burden, and in relation to the necessary (but temporary) truncation of the support of the discrete Weibull random variable. Inferential issues for the proposed model are also discussed and are eventually applied to a dataset taken from the literature, which shows that the proposed multivariate model can satisfactorily fit real-life correlated counts even better than the most popular or recent existing ones.

1. Introduction

Stochastic simulation has been playing a more and more important role in statistical research in recent years. Thanks to the increasing availability of computational resources, the evaluation of the performance of techniques of statistical analysis and the assessment of the reliability of stochastic models are now often carried out via computer-simulated data; this is the unique viable solution when handling complex estimators in inferential problems, whose statistical properties cannot be derived analytically. The researcher is often concerned with multivariate random variables (r.v.'s). In particular, multivariate discrete data or, more simply, correlated count data often arise in many contexts (statistical quality control, biostatistics, failure analysis, etc.). Such data are often modelled through the multivariate normal distribution, which, however, being a continuous r.v., fits the data hardly adequately; or through a multivariate Poisson model, which would require the data to have marginal means almost equal to the marginal variances. Here, we consider the discrete Weibull distribution as an alternative to the popular Poisson r.v. and propose a procedure for simulating correlated discrete Weibull r.v.'s, with marginal distributions and correlation matrix set by the user. The procedure really relies upon the Gaussian copula model and an iterative algorithm for recovering the proper correlation matrix for the copula ensuring

*. *Tel.: +390250321533; Fax: +3902503111*
E-mail address: alessandro.barbiero@unimi.it

doi: 10.24200/sci.2017.4412

the desired correlation matrix on the discrete margins. The rest of the paper is structured as follows: In the next section, the discrete Weibull distribution is introduced and its features are briefly described. Section 3 outlines the simulation procedure for generating correlated discrete Weibull r.v.'s with assigned margins and correlation matrix. Section 4 shows the results of a Monte Carlo simulation study, proving the performance of the proposed method, with a special focus on its capability to recover the desired correlations and its computation time. Section 5 is devoted to parameter estimation and Section 6 considers a real dataset that can be fitted by the bivariate discrete Weibull model. The final section concludes the paper with some remarks and future research perspective.

2. The discrete Weibull distribution

The discrete Weibull distribution was introduced by Nakagawa and Osaki [1] as a discrete counterpart of the continuous Weibull distribution, which was a popular stochastic model used especially in engineering reliability and survival analysis. It is also called type I discrete Weibull distribution, in order to distinguish it from other models proposed later by Stein and Dattero [2] (type II discrete Weibull) and Padgett and Spurrier [3] (type III discrete Weibull). For the model proposed by Nakagawa and Osaki [1], the cumulative distribution function (c.d.f.) is:

$$F(x; q, \beta) = 1 - q^{(x+1)^\beta}; \qquad x \in \mathbb{N}_0,$$

with $0 < q < 1$ and $\beta > 0$; consequently, its probability mass function (p.m.f.) is defined as:

$$p(x; q, \beta) = q^{x^\beta} - q^{(x+1)^\beta} \qquad x \in \mathbb{N}_0. \tag{1}$$

This distribution, differently from the two alternative types, retains the form of the c.d.f. of the continuous Weibull model. Note that for each choice of the parameter β, it results in $p(0; q, \beta) = 1 - q$, and for $\beta = 1$, the discrete Weibull r.v. reduces to the geometric r.v. with parameter $1 - q$.

The discrete Weibull distribution can be used in reliability problems for modelling discrete failure data, such as the number of shocks, cycles, or runs a component or structure can overcome before failing, or for modelling discrete lifetimes, i.e. when the lifetime of a device or a system is not measured in terms of the calendar time, but in terms of the number of periods (e.g., days, weeks, etc.) it successfully works until failure. More generally, it can virtually model any type of count data. Contrary to the Poisson r.v., which cannot adequately model count data whose variance exceeds the mean, something that often occurs in practice, the discrete Weibull r.v. can model both underdispersed and overdispersed data (see [4] and Table 1, which reports the expected value and the variance of the discrete Weibull distribution for several combinations of q and β). This distribution can also handle count data presenting an excess of zeros, which arise in many physical situations (see [4]).

By the way, one can note that the discrete Weibull r.v. can also handle only positive counts by simply modifying its c.d.f. as follows:

$$F(x; q, \beta) = 1 - q^{x^\beta}; \qquad x \in \mathbb{N},$$

whose associated p.m.f. then becomes:

$$p(x; q, \beta) = q^{(x-1)^\beta} - q^{x^\beta}; \qquad x \in \mathbb{N}.$$

With regard to the issues related to point and interval estimation of the parameters of the discrete Weibull distribution, one can refer to Khan and Khalique and Abouammoh [5], Kulasekera [6], and Barbiero [7].

It is worth noting that some univariate discrete distributions have been derived as modifications of the (type I) discrete Weibull r.v. or as discrete analogues of modified continuous Weibull r.v.'s, often with the intent of imposing a hazard function with some desired features. Among them, we remind the reader of the discrete additive Weibull distribution [8]; see [9] for a complete review.

The discrete Weibull model is implemented in the `R` environment [10] through the package `DiscreteWeibull` [11], which comprises several functions for computing the p.m.f., the c.d.f., the quantile function, and the first and second moments, and for implementing the pseudo-random generation and sample estimation.

Table 1. Expected value (E) and variance (V) for the discrete Weibull r.v. for some combinations of its parameters.

q	E	V	E	V	E	V	E	V	E	V
	$\beta = 0.5$		$\beta = 0.75$		$\beta = 1$		$\beta = 1.5$		$\beta = 2$	
0.6	7.26	291	2.48	15.3	1.5	3.75	0.93	0.95	0.74	0.49
0.7	15.2	$1.24 \cdot 10^3$	4.25	40.2	2.33	7.78	1.30	1.53	0.98	0.68
0.8	39.7	$8.07 \cdot 10^3$	8.33	141	4.00	20.0	1.96	2.83	1.38	1.04
0.9	180	$1.62 \cdot 10^5$	23.4	$1.05 \cdot 10^3$	9.00	90.0	3.55	7.61	2.23	2.12

3. Modelling correlated discrete Weibull distributions

The building, study, and application of multivariate distributions is one of the classical fields in statistics, which still continues to be an active area of research [12]. There are several methods for constructing multivariate r.v.'s. Whereas the construction based on the definition of their joint probability mass or density function poses some difficulties and often results in practical limitations, for example, in the range of possible pairwise correlations, the specification via two independent components: 1) Marginal distributions, 2) a copula function that provides the dependence structure, is much more straightforward [13,14].

Restricted to the bidimensional case, in [15], a review on constructions of discrete bivariate distributions is given. Recognizing that "Unlike their continuous analogues, discrete bivariate distributions appear to be harder to construct," a list of existing (cluster of) methods is presented and described, among which is the "construction of discrete bivariate distributions with given marginals and correlation".

The discrete Weibull distribution was employed by Englehardt and Li [4] in a correlated random multiplicative growth model for microbial counts in water; more recently, Englehardt [16] showed that the discrete Weibull distribution could model products of autocorrelated causes, generated via copula, but no other models had been proposed for building correlated discrete Weibull r.v.'s. A vector of correlated discrete Weibull r.v.'s can be easily built resorting to copulas. Moreover, using a simulation technique implemented in the `R` environment, called `GenOrd` [17], the univariate discrete Weibull distributions can be linked together through a Gaussian copula with the desired correlations. `GenOrd` was originally conceived for modelling and simulating correlated ordinal r.v.'s, i.e., variables defined on a k-point scale (with values $1, 2, \cdots, k$, see [18]); later, it was extended to the case of discrete r.v.'s (with finite or countable support) [19].

Simulating a random vector $\mathbf{X} = (X_1, \ldots, X_k)$ of k r.v.'s with margins $F_1, \cdots, F_k$ linked together by a Gaussian copula with correlation matrix $\mathbf{R}^N$ translates into the following steps:

1. Simulate from $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{R}^N)$ a k-variate standard normal distribution with correlation matrix $\mathbf{R}^N$;
2. Compute $\mathbf{U} = (\Phi(Z_1), \cdots, \Phi(Z_k))$, where Φ is the c.d.f. of the standard normal r.v.;
3. Compute $\mathbf{X} = (F_1^{-1}(U_1), \cdots, F_k^{-1}(U_k))$, where F_i^{-1} denotes the inverse c.d.f. of X_i.

The `GenOrd` procedure incorporates an algorithm, implemented by the function `ordcont`, that is able to ensure the desired correlation matrix $\mathbf{R}$ between the k discrete margins by computing the proper corresponding correlation matrix $\mathbf{R}^N$. It is well known, in fact, that the correlations between the normal components Z_i are modified throughout the transformation process leading to the desired margins. For each pair of components (X_i, X_j), the following relationship by Cario and Nelson [20] holds:

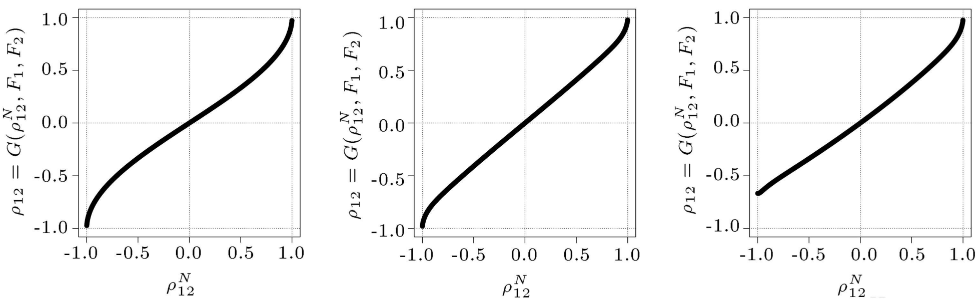
$$E(X_iX_j) = E\left[F_i^{-1}(\Phi(Z_i))F_j^{-1}(\Phi(Z_j))\right]$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_i^{-1}(\Phi(Z_i))F_j^{-1}(\Phi(Z_j)) \cdot \varphi_{\rho_{ij}^N}(z_i, z_j) dz_i dz_j, \quad (2)$$

with $\varphi_{\rho_{ij}^N}(\cdot,\cdot)$ being the standard bivariate density function with correlation coefficient ρ_{ij}^N; from Eq. (2), then, the correlation ρ_{ij} can be (usually only) numerically derived, generally different from ρ_{ij}^N. This issue is well-known in the statistical literature and several attempts to derive some properties of such "distortion" on the correlation coefficient have been made (see, for example, [20-23].

This algorithm relies upon the discretization of normal r.v.'s that necessarily requires a finite support for the target discrete r.v.'s X_i (see [18] for details). In this case, it can be shown that the double integral in Eq. (2) reduces to a finite sum of double integrals of the bivariate normal c.d.f. with correlation ρ_{ij}^N computed over rectangles in $\mathbb{R}^2$; thus, the value of ρ_{ij} can be easily derived analytically or numerically, say, $\rho_{ij} = G(\rho_{ij}^N, F_i, F_j)$, where $G(\cdot)$ is some function. This task can be worked out thanks to the availability of statistical software (`mvtnorm`) implementing the bivariate normal c.d.f. [24]. The algorithm, however, can be adapted to the case of discrete r.v.'s with countable support by operating a preliminary truncation on it; this translates into selecting a proper right bound (e.g., the $(1-\gamma)$-quantile, with $0 < \gamma << 1$ being a 'truncation parameter' [19,25]).

Figure 1 displays the relationship between ρ^N and ρ when both margins X_1 and X_2 are Bernoulli variables with probability of success 1/2 (Figure 1(a); in this case, there exists an analytical expression for function G: $\rho = 2/\pi \arcsin(\rho^N)$); or when X_1 and X_2 are Binomial variables with parameters $n = 2$ and probability of success 1/2 (Figure 1(b)); or, finally, when X_1 and X_2 are Binomial variables with parameters $n = 2$ and probability of success 3/4 (Figure 1(c)). Note also that once the two marginal distributions are assigned, the target ρ cannot be set at will in the usual range $(-1, +1)$, but has to satisfy stricter conditions (see [18,26]); this is apparent for the case displayed in Figure 1(c); here, it can be shown that ρ cannot be smaller than $-2/3$. Lower and upper bivariate bounds can be computed by the function `corrcheck` in `GenOrd`. However, note that these bounds, for $k > 2$, define



(a) X_1 and X_2 are two Bernoulli r.v.'s with $p = 0.5$ (b) X_1 and X_2 are two binomials r.v.'s with $n = 2$ and $p = 0.5$ (c) X_1 and X_2 are two binomials r.v.'s with $n = 2$ and $p = 0.75$

Figure 1. Relationship between correlation coefficients ρ^N and ρ when the components of a bivariate normal, Z_1 and Z_2, are discretized into X_1 and X_2.

necessary, but not sufficient conditions for the existence of the joint discrete distribution.

Focusing on the simulation of correlated discrete Weibull r.v.'s, the proposed procedure can be sketched out as follows, properly rearranging the scheme seen before, used for simulating r.v.'s linked by the Gaussian copula:

1. (Preliminary support truncation) For a chosen value of the truncation parameter, γ_i, for each discrete Weibull component, X_i, $i = 1, \cdots, k$, with c.d.f. F_i, compute a truncated support and then a corresponding approximate c.d.f. F_i^*;
2. (Recovering the correlation matrix of the Gaussian copula) By using the `ordcont` function in `GenOrd`, recover, through an iterative search, the correlation matrix $\mathbf{R}^N$ of a k-variate standard normal r.v. imposing the desired correlation matrix $\mathbf{R}$ on the k discrete Weibull r.v.'s with c.d.f. F_i^*;
3. (Drawing the sample) By using the `ordsample` function in `GenOrd`, draw a sample of chosen size n, $\mathbf{z} = [z_{hi}]$, $h = 1, \cdots, n$, from a k-variate standard normal r.v. with correlation matrix $\mathbf{R}^N$, apply first the standard normal c.d.f., $u_{hi} = \Phi(z_{hi})$, and then the inverse c.d.f. F_i^{-1}, $x_{hi} = F_i^{-1}(u_{hi})$. $\mathbf{x} = [x_{hi}]$ is a sample from the correlated discrete Weibull r.v.'s with c.d.f. F_i and correlation matrix $\mathbf{R}$.

4. A Monte Carlo simulation study

Through a Monte Carlo simulation study, we want to assess the capability of the proposed simulation technique of recovering the desired correlation coefficients between the discrete Weibull components and in terms of the truncation parameter, γ. We do not need to check its capability of correctly simulating each marginal distribution, since this is ensured by the 'inverse transform sampling' performed on the basis of the actual inverse c.d.f. (see [19]). Also, we want to check the computation time required by the procedure, in particular, when the dimensionality k increases. Preliminary results are reported in Barbiero [27].

4.1. Checking the "accurateness" of the procedure

In the simulation study, for the sake of simplicity, we first start focusing on the generation of $k = 3$ correlated discrete Weibull r.v.'s under various experimental settings. Specifically, we consider two sub-studies. The first one considers $k = 3$ identically distributed and equally correlated discrete Weibull r.v.'s, i.e. the simplest kind of scenarios. The three marginal distributions are all characterized by one of the following three vectors of parameters (q, β), $v_1 = (0.7, 0.75)$, $v_2 = (0.8, 1.5)$, and $v_3 = (0.9, 2)$; the r.v.'s are correlated through a common correlation coefficient ρ taking the values -0.2, 0.2, 0.4, and 0.6. Thus, 12 simulation settings arise. A negative value for the common correlation has been introduced, since, even if less often, count data in practical applications can be negatively associated (let us think about purchases of complementary products or of substitutes). The second sub-study considers three different vectors of parameters, v_1, v_2, and v_3, for the three discrete Weibull distributions X_1, X_2, and X_3, respectively; the correlation matrix $\mathbf{R}$ is characterized by three different values chosen among four (-0.2, 0.2, 0.4, and 0.6) for the three distinct correlations ρ_{12}, ρ_{13}, and ρ_{23}. Thus, 24 simulation settings arise (the number of permutations of 4 elements–the four values of ρ). This sub-study aims at testing how the simulation procedure handles more complicated scenarios than those evaluated in the first sub-study, characterized by both different margins and unequal correlations.

A first issue worth analyzing prior to proceeding with the Monte Carlo study is the effect of the truncation parameter, γ. In order to yield accurate results on the value of the target correlation coefficient, ρ, one should try to keep γ as small as possible; but a too small value of γ would entail a large number of points in the support of the (temporarily truncated) r.v., resulting in a long computational time of the `ordcont` routine of the `GenOrd` procedure. A compromise has then to be sought between precision and machine time; thus, a preliminary study can be carried out to identify an "optimal" value of γ. We performed it on the scenarios of the two simulation studies. To this aim, we considered 6 values of γ, from 10^{-2} to 10^{-7}, with a step ratio of 1/10; and, for each one of the settings, via `ordcont`, we computed the value of the common ρ^N as a function of the target ρ and of the parameter vector v. The results for the first sub-study, concerning $k = 3$ equally distributed and correlated discrete Weibull r.v.'s., are graphically displayed in Figure 2. They are interesting and somehow surprising since they show that the trend of the common ρ^N as a function of γ, keeping ρ and v fixed, is not necessarily monotonic; however, ρ^N actually reaches stability quite fast, even according to a different rate mostly depending on the combination of the values of q and β. Large values of mean and variance tend to slow down the stabilization process. For example, for the scenario characterized by $\rho = 0.2$, $q = 0.7$, and $\beta = 0.75$ for all the three random components (panel d), the common correlation coefficient ρ^N of the Gaussian copula recovered by `GenOrd` is $\rho^N = 0.2552062$ with $\gamma = 10^{-2}$, $\rho^N = 0.2633874$ with $\gamma = 10^{-3}$, $\rho^N = 0.2655006$ with $\gamma = 10^{-4}$, $\rho^N = 0.2659209$ with $\gamma = 10^{-5}$, $\rho^N = 0.2659926$ with $\gamma = 10^{-6}$, and $\rho^N = 0.2660024$ with $\gamma = 10^{-7}$. The correlation coefficient of the 3-variate standard normal r.v. is then quite different from the

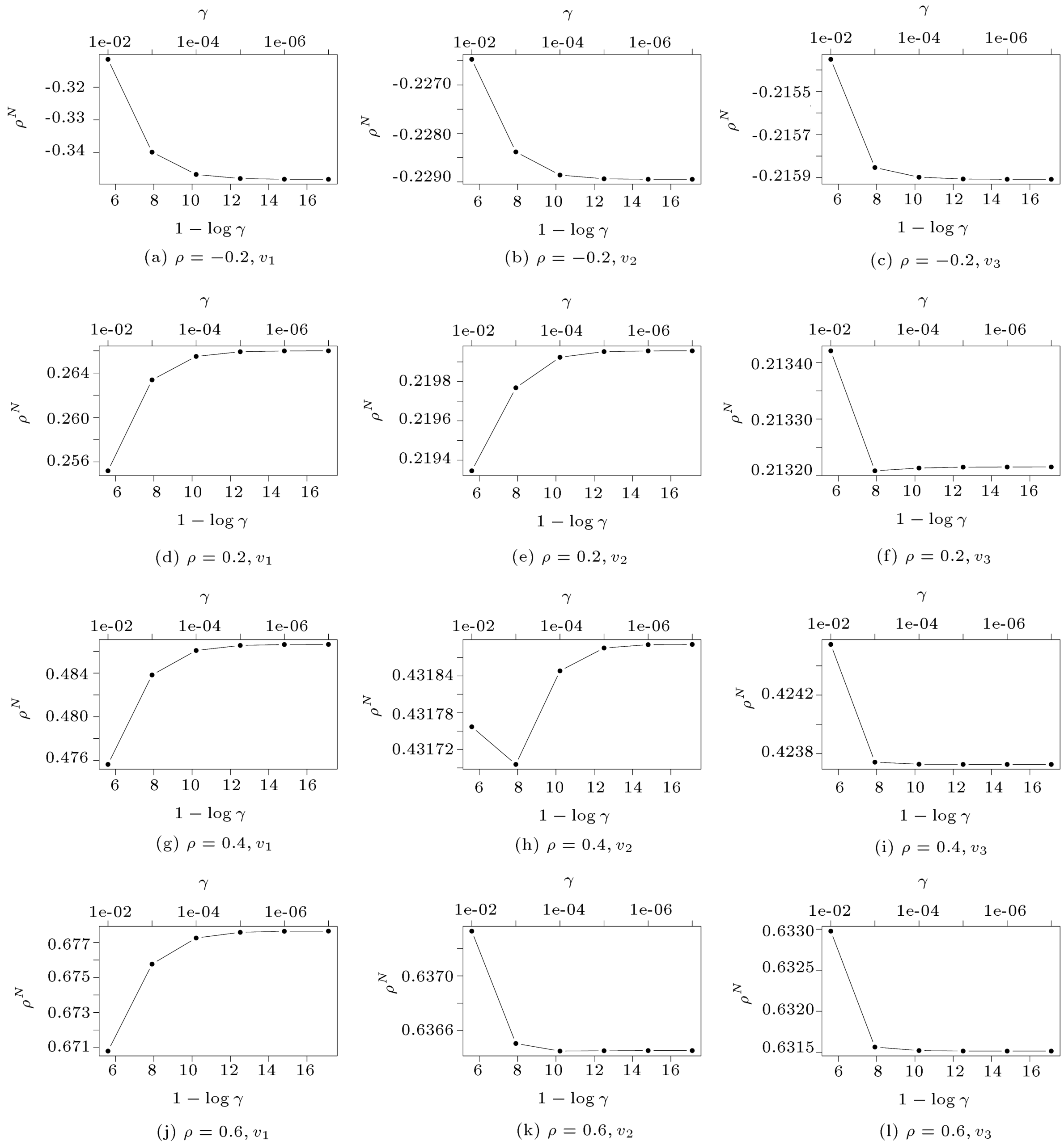


Figure 2. Value of ρ^N computed by `GenOrd` under various settings (identified by the common correlation ρ and vector of parameters v) and for different values of the truncation parameter γ.

target correlation among the discrete Weibull r.v.'s. As to the effect of the truncation parameter γ, whereas there is a non-negligible change in ρ^N moving from $\gamma = 10^{-2}$ to $\gamma = 10^{-3}$ and $\gamma = 10^{-4}$, moving from 10^{-4} to 10^{-7} yields a relative change (increase) in the correlation coefficient ρ^N of 0.19% only. Similar considerations hold for the second sub-study. For instance, for the scenario characterized by $\rho_{12} = 0.2$, $\rho_{13} = 0.4$, and $\rho_{23} = 0.6$, the correlation matrix of the Gaussian copula computed by GenOrd, according to three different values of γ ($\gamma_1 = 10^{-4}$, $\gamma_2 = 10^{-5}$, $\gamma_3 = 10^{-6}$), is:

$$\mathbf{R}^N_{\gamma_1} = \begin{pmatrix} 1 & .2462291 & .4799779 \\ .2462291 & 1 & .6370234 \\ .4799779 & .6370234 & 1 \end{pmatrix},$$

$$\mathbf{R}^N_{\gamma_2} = \begin{pmatrix} 1 & .2464809 & .4804511 \\ .2464809 & 1 & .6370470 \\ .4804511 & .6370470 & 1 \end{pmatrix},$$

$$\mathbf{R}^N_{\gamma_3} = \begin{pmatrix} 1 & .2465235 & .4805311 \\ .2465235 & 1 & .6370500 \\ .4805311 & .6370500 & 1 \end{pmatrix},$$

and is considerably different from the target $\mathbf{R}$. Also, in this case, reducing the magnitude of the truncation parameter (from 10^{-4} to 10^{-6} and beyond) does not entail a significant improvement in terms of precision of the correlation coefficients, ρ^N_{ij}, (for the three values considered here, always smaller than 0.1%). However, generally speaking, setting γ may require more caution when handling discrete Weibull marginals with large mean and variance, since a too crude truncation of the support may impose gross approximation errors on the computation of $\mathbf{R}^N$ and, on the contrary, setting a very small value of γ may dramatically increase the computation time required for recovering $\mathbf{R}^N$.

Now, $\gamma = 10^{-4}$ seems to be, at least for all the scenarios considered here, a satisfactory compromise; we decided to keep it fixed for the complete Monte Carlo study. Then, under each setting of the two sub-studies, we generate 5,000 random samples of sizes 50×3 and 100×3 through the GenOrd procedure. In order to assess its performance, we compute the Monte Carlo distribution of the three sample correlation coefficients $\hat{\rho}_{12}$, $\hat{\rho}_{13}$, and $\hat{\rho}_{23}$, which, for the values of n examined here, are nearly unbiased estimators of the corresponding ρ_{ij}. Thus, the closer the Monte Carlo mean of $\hat{\rho}_{ij}$ to the target value ρ_{ij}, the better the performance of the simulation technique.

The synthetic results are reported in Table 2 and confirm that GenOrd is able to reproduce the target correlations satisfactorily under each setting. The largest absolute "bias" (i.e., difference between target and Monte Carlo average values) for the first study is about 0.018, for the scenario with $\rho = -0.2$, v_1, and $n = 50$; whereas for the second study, it is approximately equal to 0.01 for ρ_{13} in the settings characterized by $\rho_{12} = 0.4$, $\rho_{13} = 0.6$, and $\rho_{23} = \mp 0.2$, $n = 50$. As expected, the overall bias decreases when moving from $n = 50$ to $n = 100$; thus, this study implicitly represents an assessment of the behaviour of the sample correlation as an estimator of ρ, too. Please note that for the first sub-study, for symmetry reasons, the three sample correlations should theoretically have the same sample distribution; discrepancies among the three average values under each scenario are due only to the Monte Carlo approximation. These results confirm that the preliminary truncation of the support of the discrete Weibull r.v.'s (via the parameter γ here set equal to 10^{-4}) does not affect the computation of the correlation matrix $\mathbf{R}^N$ negatively (i.e., it does not introduce sensible unwanted bias), at least for the examined settings.

As for the computation time, carrying out the simulation task takes different times depending on the specific simulation setting, more specifically on the marginal distributions. The most time-consuming settings are those involving the parameter vector ($q = 0.7, \beta = 0.75$), which determines the largest mean and variance, and then a truncated support comprising a large number of points; this leads to a longer computational time in the discretization process required for recovering the proper correlation matrix of the Gaussian copula. Nevertheless, even for the 'worst' scenario (constant correlation $\rho = 0.6$ and $\gamma = 10^{-6}$), the whole computation time is never larger than 3 minutes, the actual drawing of samples (of size 100) requiring less than 15 seconds (on a machine with Intel Core i3-2100 CPU @ 3.10 GHz, 4 GB RAM).

A remark should be made about possible problems with specific combinations of parameters. When handling small (much smaller than 1) values of β, especially if combined with values of the parameter q close to 1, which give rise to a pronounced skewness of the distribution, the discretization step required for recovering $\mathbf{R}^N$ may be computationally cumbersome, since the upper bound of the truncated support becomes very large. For 'moderate' values of the truncation parameter γ (say $0.001 \div 0.01$), the algorithm can easily recover the correlation matrix of the Gaussian copula. For very small values of γ (say $< 10^{-3}$), such task may become practically unattainable (or, better, it requires many minutes to be carried out). However, many of these (q, β) combinations produce 'unlikely' discrete Weibull distributions. Just to get an idea, if $q = 0.8$ and $\beta = 0.2$, the 0.99-quantile is $\approx 3.7 \cdot 10^6$, the expected value $\approx 2.2 \cdot 10^5$, and the standard deviation $\approx 3.4 \cdot 10^6$. However, in order to overcome this issue, alternative ways to recover the correlation matrix of the multivariate Gaussian copula

Table 2. Synthetic results of the first Monte Carlo simulation study, $k = 3$.

(a) Monte Carlo averages of sample correlations for the 12 settings of sub-study 1 ($\gamma = 10^{-4}$)

n	$\rho = -0.2$			$\rho = 0.2$			$\rho = 0.4$			$\rho = 0.6$		
50	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$
v_1	-0,218	-0,214	-0,214	0,198	0,203	0,201	0,401	0,407	0,399	0,599	0,600	0,599
v_2	-0,204	-0,198	-0,197	0,198	0,199	0,198	0,396	0,396	0,393	0,595	0,594	0,593
v_3	-0,203	-0,198	-0,196	0,198	0,198	0,196	0,396	0,397	0,394	0,595	0,595	0,595
100	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$
v_1	-0,209	-0,209	-0,209	0,203	0,203	0,201	0,398	0,403	0,401	0,597	0,600	0,599
v_2	-0,200	-0,199	-0,199	0,198	0,200	0,198	0,397	0,398	0,398	0,597	0,597	0,598
v_3	-0,200	-0,197	-0,199	0,199	0,198	0,199	0,397	0,398	0,398	0,597	0,598	0,597

(b) Monte Carlo averages of sample correlations for the 24 settings of sub-study 2 (4 scenarios for each row; values within brackets are the target correlations) ($\gamma = 10^{-4}$)

	(ρ_{12})	(ρ_{13})	(ρ_{23})	(ρ_{12})	(ρ_{13})	(ρ_{23})	(ρ_{12})	(ρ_{13})	(ρ_{23})	(ρ_{12})	(ρ_{13})	(ρ_{23})
	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$
n	(-0.2)	(0.2)	(0.4)	(-0.2)	(0.2)	(0.6)	(-0.2)	(0.4)	(0.2)	(-0.2)	(0.4)	(0.6)
50	-0.204	0.207	0.396	-0.203	0.207	0.595	-0.204	0.406	0.204	-0.205	0.408	0.594
100	-0.202	0.202	0.397	-0.202	0.202	0.596	-0.202	0.402	0.200	-0.204	0.404	0.597
n	(-0.2)	(0.6)	(0.2)	(-0.2)	(0.6)	(0.4)	(0.2)	(-0.2)	(0.4)	(0.2)	(-0.2)	(0.6)
50	-0.202	0.608	0.204	-0.206	0.607	0.400	0.205	-0.207	0.396	0.207	-0.203	0.595
100	-0.202	0.605	0.201	-0.203	0.605	0.400	0.200	-0.205	0.397	0.202	-0.202	0.598
n	(0.2)	(0.4)	(-0.2)	(0.2)	(0.4)	(0.6)	(0.2)	(0.6)	(-0.2)	(0.2)	(0.6)	(0.4)
50	0.201	0.405	-0.202	0.198	0.404	0.595	0.200	0.608	-0.203	0.200	0.606	0.396
100	0.201	0.403	-0.199	0.201	0.402	0.598	0.201	0.605	-0.200	0.202	0.604	0.398
n	(0.4)	(-0.2)	(0.2)	(0.4)	(-0.2)	(0.6)	(0.4)	(0.2)	(-0.2)	(0.4)	(0.2)	(0.6)
50	0.401	-0.206	0.198	0.407	-0.203	0.594	0.399	0.202	-0.203	0.399	0.201	0.596
100	0.401	-0.202	0.199	0.403	-0.203	0.597	0.401	0.200	-0.200	0.403	0.202	0.598
n	(0.4)	(0.6)	(-0.2)	(0.4)	(0.6)	(0.2)	(0.6)	(-0.2)	(0.2)	(0.6)	(-0.2)	(0.4)
50	0.397	0.610	-0.204	0.398	0.610	0.193	0.605	-0.201	0.204	0.603	-0.210	0.394
100	0.401	0.606	-0.200	0.399	0.606	0.199	0.604	-0.202	0.200	0.603	-0.203	0.398
n	(0.6)	(0.2)	(-0.2)	(0.6)	(0.2)	(0.4)	(0.6)	(0.4)	(-0.2)	(0.6)	(0.4)	(0.2)
50	0.603	0.202	-0.203	0.604	0.200	0.393	0.603	0.403	-0.204	0.603	0.404	0.194
100	0.602	0.201	-0.201	0.603	0.202	0.398	0.602	0.403	-0.201	0.602	0.403	0.198

that induces the target correlation matrix, i.e. handling Eq. (2), can be inspected, for example resorting to some approximation by the continuous Weibull model.

4.2. Checking the "scalability" of the procedure

In a second empirical study, we try to address the "scalability" of the algorithm, that is, to check for its tightness when increasing the dimensionality k of the random vector to be simulated. Here, we first consider $k = 6$ discrete Weibull components, whose margins are characterized by the following q and β parameters: $\mathbf{q} = (q_1, \cdots, q_6) = (0.7, 0.8, 0.8, 0.8, 0.8, 0.9)$ and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_6) = (0.75, 0.75, 1, 1.5, 2, 2)$, and whose correlations are set equal to a constant value ρ, with $\rho = -0.1, 0.2, 0.4, 0.6$. Then, we consider the simulation of $k = 20$ correlated discrete Weibull r.v.'s with marginal parameters given by $\mathbf{q} = (q_1, \cdots, q_{20}) = (.7, .7, .7, .7, .7, .7, .7, .7, .8, .8, .8, .8, .8, .8, .8, .8, .9, .9, .9,$

.9) and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{20}) = (.75, .75, 1, 1, 1.5, 1.5, 2, 2, .75, .75, 1, 1, 1.5, 1.5, 2, 2, 1.5, 1.5, 2, 2)$, and the same level of (constant) correlation ($\rho = 0.2, 0.4, 0.6$). Under each setting of the two sub-studies, we generate 5,000 random samples of size $100 \times k$ through the `GenOrd` procedure. The value of γ is varied (10^{-4}, 10^{-5}, 10^{-6}). The performance of the algorithm is still synthetically assessed through the MC average of the sample correlations $\hat{\rho}_{ij}$. The total computation time is measured as the sum of two distinct contributions: the first is the time required by the `ordcont` procedure for setting up the correlation matrix $\mathbf{R}^N$ of the k-variate normal r.v., which we call "preprocessing time", t_1; the second is the time requested by the actual simulation of 5,000 multivariate samples, t_2.

The results for $k = 6$ and $\gamma = 10^{-4}$ are reported in Table 3; the MC average value of the sample correlations $\hat{\rho}_{ij}$ is always close to the corresponding ρ_{ij}, for any choice of the target ρ_{ij} and for each choice of the pair (i, j).

In Tables 4 and 5, the computation times for each of the explored settings are reported for $k = 6$ and $k = 20$, respectively. For the same experimental scenario, an essential role in determining the time t_1 is played by the truncation parameter, γ; smaller values of γ lead to larger values of t_1. Moreover, the value of t_1 is influenced by the level of correlation among the discrete Weibull components once the value of the truncation parameter, γ, is fixed. Larger values of ρ (in absolute value) require a larger preprocessing time t_1 (this means that `ordcont` needs more iterations for convergence when $|\rho|$ moves to 1).

In greater detail, with $k = 6$ correlated discrete Weibull r.v.'s, the pre-processing time t_1 is about 1 minute or less, setting $\gamma = 10^{-4}$, whereas, when choosing $\gamma = 10^{-6}$, it grows up till 3 minutes; the largest value for t_1 is recorded with $\rho = 0.6$ (the convergence of the iterative procedure implemented by the `R` function `ordcont` requires a greater number of steps).

When $k = 20$, for the explored scenarios, it takes up to 9 minutes for recovering the correct correlation matrix $\mathbf{R}^N$ of the multivariate normal with $\gamma = 10^{-4}$; this time grows up till 23 minutes, when setting $\gamma = 10^{-6}$. The worst scenario is again that characterized by the larger absolute value of ρ (0.6).

Summarizing these results, we can state that the `GenOrd` procedure is not negatively affected by the increasing dimensionality k; naturally, the preprocessing computation time increases with a quadratic rate in k (finding the correct $\mathbf{R}^N$ consists of $k(k-1)/2$ iterative searches), but there are no further obstacles making the simulation more time-demanding.

Table 4. Computation time in seconds (for the preprocessing step, t_1, and for the simulation of 5,000 samples of size $n = 100$, t_2) for $k = 6$ correlated discrete Weibull r.v.'s.

ρ	$\gamma = 10^{-4}$		$\gamma = 10^{-5}$		$\gamma = 10^{-6}$	
	t_1	t_2	t_1	t_2	t_1	t_2
−0.1	47.8	24.8	67.2	29.0	116.8	30.8
0.2	38.0	23.6	92.2	33.4	152.7	32.9
0.4	60.3	25.9	108.6	29.0	165.4	27.4
0.6	65.5	23.4	134.3	28.3	188.5	30.1

Table 5. Computation time in minutes (for the preprocessing step, t_1, and for the simulation of 5,000 samples of size $n = 100$, t_2) for $k = 20$ correlated discrete Weibull r.v.'s.

ρ	$\gamma = 10^{-4}$		$\gamma = 10^{-5}$		$\gamma = 10^{-6}$	
	t_1	t_2	t_1	t_2	t_1	t_2
0.2	6.07	1.25	9.69	1.19	13.99	1.14
0.4	7.52	1.27	13.03	1.45	18.23	1.13
0.6	8.81	1.20	14.52	1.14	23.16	1.18

5. Inference

In order to estimate the parameters of the multivariate discrete Weibull model proposed, a two-step procedure is here suggested, which can be considered as a modification of the so called Inference Function of Margins (IFM), already present in the copula literature [28]. The first step works as the components of the r.v. were independent; the second step takes into account the association among them. The procedure works as follows:

1. Estimate the marginal c.d.f.'s, F_i, $i = 1, \cdots, k$, by estimating the associated distribution parameters, θ_i, through the maximum likelihood method, thus obtaining $\hat{F}_i = F_i(\hat{\theta}_i)$;

Table 3. Results of the Monte Carlo simulation study with $k = 6$ ($\gamma = 10^4$): Monte Carlo averages of sample pairwise correlations over 5,000 samples of size $n = 100$.

ρ	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{14}$	$\hat{\rho}_{15}$	$\hat{\rho}_{16}$	$\hat{\rho}_{23}$	$\hat{\rho}_{24}$	$\hat{\rho}_{25}$	$\hat{\rho}_{26}$	$\hat{\rho}_{34}$	$\hat{\rho}_{35}$	$\hat{\rho}_{36}$	$\hat{\rho}_{45}$	$\hat{\rho}_{46}$	$\hat{\rho}_{56}$
-.1	-.102	-.103	-.098	-.102	-.100	-.103	-.103	-.102	-.100	-.101	-.100	-.100	-.100	-.101	-.100
.2	.202	.202	.203	.202	.203	.203	.200	.202	.202	.198	.202	.202	.201	.197	.199
.4	.402	.400	.405	.402	.403	.400	.403	.404	.405	.400	.401	.399	.400	.399	.398
.6	.601	.602	.604	.606	.605	.602	.603	.606	.606	.599	.601	.601	.599	.597	.598

2. • Estimate the correlation coefficient ρ_{ij}, for each $1 \leq i < j \leq k$, e.g., through the sample correlation coefficient $\hat{\rho}_{ij}$;
   • Using `ordcont` in `GenOrd`, numerically compute $\hat{\rho}^N_{ij} = G^{-1}(\hat{\rho}_{ij}, \hat{F}_i, \hat{F}_j)$, for each $1 \leq i < j \leq k$, and thus reconstruct the Gaussian copula correlation matrix $\hat{\mathbf{R}}^N = [\hat{\rho}^N_{ij}]$.

Note that here the number of parameters to be estimated is $2k$ (two parameters for each discrete Weibull marginal) plus $k(k-1)/2$ (the distinct pairwise correlation coefficients of the Gaussian copula) for a total of $k(k+3)/2$.

This two-step procedure for the estimation of parameters is much more straightforward to apply (and more intuitive and easy to understand) than the alternative procedure that maximizes the log-likelihood ℓ of the model with respect to all the parameters simultaneously. This is particularly true when the dimensionality k increases. In fact, the log-likelihood for the copula based discrete Weibull model can be written as:

$$\ell(\mathbf{q}, \boldsymbol{\beta}, \mathbf{R}^N; \mathbf{x}) = \sum_{i=1}^{n} \log P\left(x_{i1}, \cdots, x_{ik}; \mathbf{q}, \boldsymbol{\beta}, \mathbf{R}^N\right), \tag{3}$$

where the probability $P(x_{i1}, \cdots, x_{ik}; \mathbf{q}, \boldsymbol{\beta}, \mathbf{R}^N)$ is obtained as a rectangle probability of the underlying multivariate normal distribution, i.e. as a multiple integral involving the k-variate standard normal density function $\phi_k(z_1, \cdots, z_k)$ only:

$$P(x_{i1}, \cdots, x_{ik}; \mathbf{q}, \beta, \mathbf{R}^N) = \int_{\Phi^{-1}[F_1(x_{i1}-1;q_1,\beta_1)]}^{\Phi^{-1}[F_1(x_{i1};q_1,\beta_1)]} \cdots \int_{\Phi^{-1}[F_k(x_{ik}-1;q_k,\beta_k)]}^{\Phi^{-1}[F_k(x_{ik};q_k,\beta_k)]} \phi_k(z_1, \cdots, z_k; \mathbf{R}^N) dz_1 \cdots dz_k.$$

Thus, deriving parameter estimates for the Weibull model by directly maximizing the joint log-likelihood in Eq. (3) can be a very hard task, entailing a severe computational burden [14].

6. An application to a real dataset

In this section, we apply the proposed multivariate discrete Weibull distribution to model a dataset taken from the literature. The data, considered in [29] (see Table 6), consist of the number of aborts by 109 aircrafts in two (first = x_1, second = x_2) consecutive 6 months of a 1-year period. They can be regarded as repeated observations on the same unit, thus, leading to a panel data.

Table 6. Bivariate distribution of the data taken from [29]; number of flight aborts by 109 aircrafts in the first and second consecutive six months of a one-year period.

$x_1 \setminus x_2$	0	1	2	3	4	
0	34	20	4	6	4	68
1	17	7	0	0	0	24
2	6	4	1	0	0	11
3	0	4	0	0	0	4
4	0	0	0	0	0	0
5	2	0	0	0	0	2
	59	35	5	6	4	109

Mitchell and Paulson [29] used a new bivariate negative binomial distribution derived by convoluting a bivariate geometric distribution to model the data.

In order to fit the proposed bivariate discrete Weibull model to these data, we first compute the maximum likelihood estimates for the parameters of the two margins X_1 and X_2 (the computation can be carried out by using the function `estdweibull` within the `DiscreteWeibull` package, with the option `method=''ML''`). We obtain $\hat{q}_1 = 0.3788, \hat{\beta}_1 = 0.9774$, $\hat{q}_2 = 0.4496, \hat{\beta}_2 = 1.1202$, with standard errors $se(\hat{q}_1) = 0.0459$, $se(\hat{\beta}_1) = 0.1177$, $se(\hat{q}_2) = 0.0464$ and $se(\hat{\beta}_2) = 0.1204$. Note that both the β parameters are quite close to 1 (the first is slightly smaller, the second a bit larger); this confirms that both marginal distributions are "close" to geometric and somewhat justifies the use of a bivariate negative binomial distribution by Mitchell and Paulson [29].

Then, we compute the sample correlation coefficient between x_1 and x_2, resulting in $\hat{\rho} = -0.1608955$ from which, by using the function `ordcont` in `GenOrd` (again with $\gamma = 0.0001$) with the maximum likelihood estimates computed above, we derive an estimate of the correlation coefficient of the Gaussian copula: $\hat{\rho}^N = -0.2588228$.

From all these five parameter estimates, we can then easily derive the (estimated) theoretical joint p.m.f. by particularizing what was described in the previous section for the case of $k = 2$. For example, in order to calculate the joint probability in $(0, 0)$, we compute:

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= \Phi_2\left[\Phi^{-1}(F_1(0; \hat{q}_1, \hat{\beta}_1)), \Phi^{-1}(F_2(0; \hat{q}_2, \hat{\beta}_2)); \rho^N = -0.2588228\right] \\ &= \Phi_2\left[\Phi^{-1}(1-\hat{q}_1), \Phi^{-1}(1-\hat{q}_2), \rho^N = -0.2588228\right] \\ &= 0.3027162. \end{aligned}$$

More generally, after defining:

$$C_{\rho^N}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \phi(s,t;\rho^N) ds dt,$$

the joint c.d.f. of (X_1, X_2) is given by:

$$F(x_1,x_2) = C_{\rho^N}(F_1(x_1), F_2(x_2)),$$

and then, for each possible $(i,j) \in \mathbb{N}_0^2$, we have:

$$\begin{aligned} P(X_1=i, X_2=j) = & C_{\rho^N}\left[F_1\left(i;\hat{q}_1,\hat{\beta}_1\right), F_2\left(j;\hat{q}_2,\hat{\beta}_2\right)\right] \\ & - C_{\rho^N}\left[F_1\left(i-1;\hat{q}_1,\hat{\beta}_1\right), F_2\left(j;\hat{q}_2,\hat{\beta}_2\right)\right] \\ & - C_{\rho^N}\left[F_1\left(i;\hat{q}_1,\hat{\beta}_1\right), F_2\left(j-1;\hat{q}_2,\hat{\beta}_2\right)\right] \\ & + C_{\rho^N}\left[F_1\left(i-1;\hat{q}_1,\hat{\beta}_1\right), F_2\left(j-1;\hat{q}_2,\hat{\beta}_2\right)\right], \end{aligned}$$

letting $F_i(-1;q,\beta) = 0$, $i = 1,2$.

For the dataset at hand, the theoretical joint p.m.f. is reported in Table 7 whereas the expected joint frequencies are displayed in Table 8.

To test the null hypothesis H_0: "The data come from the bivariate discrete Weibull model" against the alternative H_1: "The data do not come from the bivariate discrete Weibull model", we employ the asymptotic chi-squared goodness-of-fit test, based on the statistic:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*},$$

where n_{ij} is the observed frequency of $(i,j) \in \mathbb{N}_0^2$ and n_{ij}^* is its theoretical analogue. Since $n = 109$, it can be usefully employed for assessing whether the bivariate discrete Weibull model fits the data adequately. Some problems arise on how to collapse adjacent cells, to ensure every expected frequency is greater than 5 (in order to apply the asymptotic results satisfactorily).

Table 8. Theoretical joint absolute frequencies for the data from [29] under the bivariate discrete Weibull model.

$x_1 \setminus x_2$	0	1	2	3	≥ 4	
0	33.00	20.13	8.94	3.57	2.07	67.71
1	15.59	6.35	2.19	0.71	0.32	25.16
2	6.65	2.17	0.66	0.20	0.08	9.76
3	2.78	0.76	0.21	0.06	0.02	3.83
4	1.15	0.27	0.07	0.02	0.01	1.52
≥ 5	0.81	0.15	0.04	0.01	0.00	1.01
	59.99	29.83	12.11	4.57	2.50	109

For the grouping displayed in Table 9(a), the value of the statistic is $\chi^2 = 9.573$. The χ^2 statistic under H_0 (under our model) is asymptotically distributed as an χ^2 r.v. with 9-5-1 degrees of freedom (the number of classes being 9; the number of parameters 5). The corresponding p-value takes the value 0.0226 and indicates quite a bad fit of the model.

Moving to the grouping displayed in Table 9(b), the value of the statistic is $\chi^2 = 3.807$. The χ^2 statistic under H_0 (under our model) is distributed as an χ^2 r.v. with 8-5-1 degrees of freedom and the corresponding p-value takes the value 0.149, indicating that we accept the hypothesis that the data come from the bivariate discrete Weibull model at any significance level smaller than or equal to 0.149.

If we compute the log-likelihood:

$$\ell = \sum_i \sum_j n_{ij} \log \hat{p}(i,j),$$

where $\hat{p}(i,j)$ is the estimate of the probability $P(X_1 = i, X_2 = j) = p(i,j)$ (to be adapted in case i or j are the maximum observed values of X_1 and X_2, respectively), based on the IFM estimates of the bivariate model, we obtain $\ell = -243.7517$. If we want to compare this model with some competitors, we can resort to the Akaike Information Criterion (AIC), given by AIC $= 2\kappa - 2\hat{\ell}$, where κ is the number of parameters of the model and $\hat{\ell}$ the maximum value of the log-likelihood function: the preferred model is the one with the minimum AIC value. For

Table 7. Theoretical bivariate p.m.f. for the data from Mitchell and Paulson [29] under the bivariate discrete Weibull model.

$x_1 \setminus x_2$	0	1	2	3	≥ 4	
0	0.3027	0.1846	0.0820	0.0328	0.0190	0.6212
1	0.1430	0.0583	0.0201	0.0065	0.0029	0.2309
2	0.0610	0.0199	0.0061	0.0018	0.0007	0.0895
3	0.0256	0.0070	0.0019	0.0005	0.0002	0.0352
4	0.0106	0.0025	0.0006	0.0002	0.0001	0.0139
≥ 5	0.0074	0.0014	0.0003	0.0001	0.0000	0.0093
	0.5504	0.2737	0.1111	0.0419	0.0229	1

Table 9. Sets of groupings used for computing the chi-squared statistic for the data from [29]. Contiguous cells shaded with the same colour belong to the same grouping; cells with white background colour have not been aggregated.

(a) First set of groupings

$x_1 \setminus x_2$	0	1	2	3	≥ 4
0	33.00	20.13	8.94	3.57	2.07
1	15.59	6.35	2.19	0.71	0.32
2	6.65	2.17	0.66	0.20	0.08
3	2.78	0.76	0.21	0.06	0.02
4	1.15	0.27	0.07	0.02	0.01
≥ 5	0.81	0.15	0.04	0.01	0.00

(b) Second set of groupings

$x_1 \setminus x_2$	0	1	2	3	≥ 4
0	33.00	20.13	8.94	3.57	2.07
1	15.59	6.35	2.19	0.71	0.32
2	6.65	2.17	0.66	0.20	0.08
3	2.78	0.76	0.21	0.06	0.02
4	1.15	0.27	0.07	0.02	0.01
≥ 5	0.81	0.15	0.04	0.01	0.00

the Bivariate Discrete Weibull model (BDW), we do not compute the value of the AIC exactly (which would require the simultaneous computation of all the MLEs), but we can derive an upper bound as $\text{AIC}^*_{\text{BDW}} = 2\kappa - 2\ell = 497.5034 \geq \text{AIC}_{\text{BDW}}$; for the Bivariate Negative Binomial distribution (BNB), which has six parameters, we obtain $\text{AIC}_{\text{BNB}} = 500.532$; for the Bivariate Generalized Poisson distribution (BGP) proposed by Famoye [30], which has five parameters, $\text{AIC}_{\text{BGP}} = 498.594$. Since the minimum AIC is obtained for the proposed bivariate discrete Weibull model, we can conclude it represents a good choice for the aircrafts data.

7. Conclusion

We proposed a procedure for generating correlated discrete Weibull r.v.'s with assigned marginal distributions and correlation matrix. A software implementation is provided in the `R` environment, which on the one hand simulates univariate discrete Weibull r.v.'s (package `DiscreteWeibull`) and on the other hand (`GenOrd`) links the r.v.'s together through a Gaussian copula, ensuring the desired pairwise correlations. This multivariate discrete model can handle under- and overdispersion in the data, unlike multivariate Poisson models–by far, the most popular when modelling correlated count data–which can handle equi-dispersed data only. Estimation of the model through a modification of the so-called Inference Function of Margins (IFM) is straightforward and much easier than through the maximum likelihood method. The performance of the simulation technique, specifically the capability of `GenOrd` in reproducing the desired correlations, is here briefly demonstrated through a simulation study, which considers several settings. The procedure is shown not to be time-consuming for the examined scenarios. However, some combinations of parameters of the discrete Weibull r.v., in particular those involving values of β much smaller than 1, by giving rise to very skew distributions, can critically slow down the procedure. Future research will better inspect this point and try to overcome it, possibly resorting to some approximation by the continuous Weibull model when estimating the correlation of the Gaussian copula. Future research will also further address parameter estimation and evaluation of goodness-of-fit of the model. In particular, the need for testing the Gaussian copula structure's adequacy will be of uttermost importance, since it is well known that it may fail to capture dependence between extreme events, and will require choosing through or adapting the proposals that have recently been made for goodness-of-fit testing of copula models.

References

1. Nakagawa, T. and Osaki, S. "The discrete Weibull distribution", *IEEE Transactions on Reliability*, **24**(5), pp. 300-301 (1975).

2. Stein, W.E. and Dattero, R. "A new discrete Weibull distribution", *IEEE Transactions on Reliability*, **33**(2), pp. 196-197 (1984).

3. Padgett, W.J. and Spurrier, J.D. "Discrete failure models", *IEEE Transactions on Reliability*, **34**(3), pp. 253-256 (1985).

4. Englehardt, J.D. and Li, R.C. "The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water", *Risk Analysis*, **31**(3), pp. 370-381 (2011).

5. Khan, M.S.A., Khalique, A. and Abouammoh, A.M. "On estimating parameters in a discrete Weibull distribution", *IEEE Transactions on Reliability*, **38**(3), pp. 348-350 (1989).

6. Kulasekera, K.B. "Approximate MLE's of the parameters of a discrete Weibull distribution with type I censored data", *Microelectronics Reliability*, **34**(7), pp. 1185-1188 (1984).

7. Barbiero, A. "A comparison of methods for estimating parameters of the type I discrete Weibull distribution", *Statistics and Its Interface*, **9**(2), pp. 203-212 (2016).

8. Bebbington, M., Lai, C.D., Wellington, M. and Zitikis, R. "The discrete additive Weibull distribution: A bathtub-shaped hazard for discontinuous failure data", *Reliability Engineering & System Safety*, **106**, pp. 37-44 (2012).

9. Lai, C.D. "Discrete Weibull Distributions and Their Generalizations", In *Generalized Weibull Distributions*, Springer Berlin Heidelberg, pp. 97-113 (2014).

10. R Core Team. "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria (2016). URL https://www.R-project.org/

11. Barbiero, A. "DiscreteWeibull: Discrete Weibull distributions (Type 1 and 3)", R package version 1.1. http://CRAN.R-project.org/package=Discrete Weibull (2015).

12. Sarabia, J.M. and Gòmez-Déniz, E. "Construction of multivariate distributions: A review of some recent results", *SORT*, **32**(1), pp. 3-36 (2008).

13. Nelsen, R.B., *An Introduction to Copulas*, Springer, New York (1999).

14. Nikoloulopoulos, A.K. "Copula-based models for multivariate discrete response data", *Copulae in Mathematical and Quantitative Finance, Lecture Notes in Statistics*, In P. Jaworski et al., Eds., pp. 231-249 (2013).

15. Lai, C.D. "Constructions of discrete bivariate distributions", *Advances in Distribution Theory, Order Statistics, and Inference*, in N. Balakrishnan et al., Eds., pp. 29-58 (2006). Birkhäuser Boston.

16. Englehardt, J. "Distributions of Autocorrelated First-Order Kinetic Outcomes: Illness Severity", *PLoS ONE*, **10**(6) (2015). DOI: 10.1371/journal.pone.0129042

17. Barbiero, A. and Ferrari, P.A. "GenOrd: Simulation of ordinal and discrete variables with given correlation matrix and marginal distributions", R package version 1.4.0. http://CRAN.R-project.org/package=GenOrd (2015).

18. Ferrari, P.A. and Barbiero, A., "Simulating ordinal data", *Multivariate Behavioral Research*, **47**(4), pp. 566-589 (2012).

19. Barbiero, A. and Ferrari, P.A. "Simulating correlated ordinal and discrete variables with assigned marginal distributions", in V. Melas, S. Mignani, P. Monari, L. Salmaso, Eds., *Topics in Statistical Simulation. Research from the 7th International Workshop on Statistical Simulation*, Springer-Verlag, New York (2014).

20. Cario, M.C. and Nelson, B.L. "Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix", Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois (1997).

21. Nelson, B.L. *Foundations and Methods of Stochastic Simulation: A First Course*, Springer, New York (2013).

22. Madsen, L. and Dalthorp, D. "Simulating correlated count data", *Environmental and Ecological Statistics*, **14**(2), pp. 129-148 (2007).

23. Madsen, L. and Birkes, D. "Simulating dependent discrete data", *Journal of Statistical Computation and Simulation*, **83**(4), pp. 677-691 (2013).

24. Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. and Hothorn, T. "mvtnorm: Multivariate normal and t distributions", R package version 1.0-2. http://CRAN.R-project.org/package=mvtnorm (2014).

25. Barbiero, A. and Ferrari, P.A. "Simulating correlated Poisson variables", *Applied Stochastic Models in Business and Industry*, **31**, pp. 669-680 (2015).

26. Demirtas, H. and Hedeker, D. "A practical way for computing approximate lower and upper correlation bounds", *The American Statistician*, **65**, pp. 104-109 (2011).

27. Barbiero, A. "Simulating correlated discrete Weibull variables: a proposal and an implementation in the R environment", *International Conference of Computational Methods in Science and Engineering*, Athens, 20-23 March 2015. AIP Conference Proceedings, **1702**(190017) (2015).

28. Joe, H. "Asymptotic efficiency of the two-stage estimation method for copula-based models", *Journal of Multivariate Analysis*, **94**(2), pp. 401-419 (2005).

29. Mitchell, C.R. and Paulson, A.S. "A new bivariate negative binomial distribution", *Naval Research Logistics Quarterly*, **28**, pp. 359-374 (1981).

30. Famoye, F. "A new bivariate generalized Poisson distribution", *Statistica Neerlandica*, **64**(1), pp. 112-124 (2010).

Biography

Alessandro Barbiero obtained his MS degree in Management Engineering from the University of Udine (2001) and his PhD in Statistics from the University of Milano-Bicocca (2009). He is now working as a post-doc researcher in Statistics in the Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano. His research areas include stochastic models for reliability, resampling techniques, simulation of multivariate random variables, and imputation of missing values.