Discrete analogues of continuous bivariate probability distributions

Alessandro Barbiero

Abstract In many real-world applications, the phenomena of interest are continuous in nature and modeled through continuous probability distributions, but their observed values are actually discrete and hence it would be more reasonable and convenient to choose an appropriate (multivariate) discrete distribution generated from the underlying continuous model preserving one or more important features. In this paper, two methods are discussed for deriving a bivariate discrete probability distribution from a continuous one by retaining some specific features of the original stochastic model, namely 1) the joint density function, or 2) the joint survival function. Examples of applications are presented, which involve two types of bivariate exponential distributions, in order to illustrate how the discretization procedures work and show whether and to which extent they alter the dependence structure of the original model. We also prove that some bivariate discrete distributions that were recently proposed in the literature can be actually regarded as discrete counterparts of well-known continuous models. A numerical study is presented in order to illustrate how the procedures are practically implemented and to present inferential aspects. Two real datasets, considering correlated discrete recurrence times (the former) and counts (the latter) are eventually fitted using two discrete analogues of a bivariate exponential distribution.

Keywords bivariate exponential distribution · bivariate failure rates · copula · correlated counts · linear correlation · survival function

A. Barbiero
Department of Economics, Management and Quantitative Methods,
Università degli Studi di Milano, via Conservatorio 7, 20122 Milan, Italy
Tel.: +390250321533
Fax: +3902503111
E-mail: alessandro.barbiero@unimi.it

1 Introduction

Deriving discrete analogues of continuous probability distributions has recently drawn attention of researchers; this procedure is often referred to as "discretization", even if the word may sound a bit ambiguous, being used also in other broader or different contexts (just think about discretization of ordinary, partial or stochastic differential equations or continuous time/space processes).

In many real-world applications, e.g., in reliability engineering (Bebbington et al., 2012), the random variables modeling the phenomena of interest are continuous in nature, but their observed values are actually discrete and hence it is reasonable and convenient to choose an appropriate (multivariate) discrete distribution generated from the underlying continuous model preserving one or more important features. In the context of multivariate statistical analysis, the researcher sometimes deals with observable variables that are categorized into classes or measured as discrete, and can be regarded as the result of coarse-grained measurement (discretization) of an underlying continuous (unobservable or latent) variable (Loehlin, 2004).

Finding a discrete analogue to a continuous distribution can be also useful in some applications where dealing with a continuous model is computationally cumbersome and properly substituting it with a discrete version can produce an approximate but still acceptable solution with a much smaller computational effort. In this sense, a relevant example in the engineering field is the computation of the reliability parameter in complex stress-strength models (Kotz et al., 2003). If the distributions of strength and stress are known, then the reliability can be obtained by using ordinary transformation techniques, but when the functional relationships of strength and stress are complex, such analytical techniques are intractable and alternative techniques must be adopted to arrive at a close approximation for the actual reliability. One viable technique is represented by discretization: the stress and strength continuous random components are discretized and an approximate value of the reliability parameter of the original continuous set-up is computed on the discretized set-up (Roy and Dasgupta, 2001). In stochastic programming, the evaluation of the expected value of some continuous stochastic quantity appearing in the objective function may require, among all, multivariate numerical integration, which is a rather cumbersome task; this obstacle can be overcome by approximating the continuous distribution via a finite number of points, called "scenarios", which allow to model the stochastic problem as a deterministic optimization problem (see e.g. Ruszczyński and Shapiro, 2003). In Tancrez et al. (2011), in the context of analysis of manufacturing systems with finite capacity, the authors suggest discretization of continuous processing time distributions with finite support into discrete probability distributions with a finite number of points, which result more tractable for subsequent analytical modeling through a Markov chain.

In recent decades, a large number of research papers dealing with a discrete distribution derived by discretizing a continuous one, according to some

criterion, have appeared in a scattered manner in statistical literature. There are several ways to derive a discrete distribution from a continuous one and no universally accepted criterion exists for producing a distribution that can unequivocally be called the discrete version of a continuous one. In the current literature, we can find three papers that deal with discrete analogues of continuous distributions in depth. Bracquemond and Gaudoin (2003) carried out a survey on discrete lifetime distributions derived from continuous ones; Lai (2013) described issues in the construction of such discrete lifetime distributions, and more recently, Chakraborty (2015) provided an exhaustive review of different methods of generating discrete probability distributions as analogues of continuous probability distributions.

Methods and issues related to the construction of bivariate discrete distributions, which are disseminated in the literature, have been reviewed in Lai (2006). Sixteen (clusters of) techniques have been there identified and described; however, there is no explicit mention at any "discretization" technique, capable of recovering a bivariate discrete distribution from a bivariate continuous distribution. In this paper, generalizing extant univariate techniques, we illustrate two methods that can be usefully employed in order to construct a discrete counterpart to a bivariate continuous distribution. The work is structured as follows: in the next section, we recall two methods for the discretization of univariate random variables; Section 3 introduces and describes their extension to the bivariate case highlighting their properties; Section 4 illustrates two examples of applications involving just as many bivariate continuous distributions; Section 5 provides a numerical example; in Section 6, real datasets are fitted by two discrete analogues of a bivariate exponential distribution; Section 7 concludes the paper with some final remarks.

2 Discretization methods in the univariate context

In the univariate context, a continuous random variable (rv) may be characterized either by its probability density function (pdf), cumulative distribution function (cdf), survival function (sf), moment generating function, moments, hazard rate function, etc. Basically, the construction of a discrete analogue from a continuous distribution is based on the principle of preserving one or more characteristic properties. Within this framework, two of the most popular and intuitive discretization methods are those preserving i) the pdf, ii) the sf of the continuous rv; they will be now briefly recalled.

2.1 Discretization preserving the pdf

In this method (see Chakraborty, 2015, p.4, where it is referred to as "Methodology II") a discrete rv X is derived as an analogue of the continuous rv W

with pdf $f_W(w)$, $-\infty < w < +\infty$, by defining its pmf as follows:

$$P(X = k) = f_W(k) / \sum_{j=-\infty}^{+\infty} f_W(j) \quad k \in \mathbb{Z}. \tag{1}$$

The pmf (1) may not always have a closed expression due to the infinite series sum at the denominator, which acts as a normalizing constant.

Examples of discrete distributions derived following this method are the discrete Laplace distribution by Inusah and Kozubowski (2006), the discrete skew Laplace distribution by Kozubowski and Inusah (2006); and the discrete Normal distribution proposed by Kemp (1997).

2.2 Discretization preserving the sf

This method (see Chakraborty, 2015, pp.9-10, where it is referred to as "Methodology IV") considers a continuous rv W with sf $S_W(w)$ and derives a discrete rv X whose pmf is given by:

$$P(X = k) = F_W(k+1) - F_W(k) = S_W(k) - S_W(k+1) \quad k \in \mathbb{Z}. \tag{2}$$

This discrete analogue preserves the sf, since $S_X(k) = S_W(k)$ for each integer k. The resulting pmf is in a compact form if the continuous sf $S_W(w)$ is. It can be shown that X is equal in distribution to $\lfloor W \rfloor$ (largest integer less than or equal to W).

Two examples of discrete distributions derived following this criterion are the discrete Weibull distribution introduced by Nakagawa and Osaki (1975) and the discrete normal distribution by Roy (2003).

If applied to the same continuous distribution, the two methods above generally give birth to two different discrete distributions. A very important exception is represented by the exponential distribution, with pdf $f(w) = \lambda e^{-\lambda w}$, $w > 0$, $\lambda > 0$, and sf $S_W(w) = e^{-\lambda w}$; in this case, the two methods yield the same discrete distribution X, with pmf $p(x) = e^{-\lambda x}(1 - e^{-\lambda})$, $x \in \mathbb{N}$, i.e., a geometric rv with parameter $\theta = e^{-\lambda}$ (providing the probability of non-zero values).

These two methods are open to being extended to the bivariate context; this will be discussed in the next section.

3 Extending discretization to the bivariate context

Henceforth, we will consider a bivariate continuous rv (W, Z), with joint pdf $f(w, z)$, cdf $F_{WZ}(w, z) := P(W \leq w, Z \leq z)$ and sf $S_{WZ}(w, z) := P(W \geq w, Z \geq z)$, which we want to discretize into a bivariate discrete rv (X, Y), with joint pmf $p(x, y)$, cdf $F_{XY}(x, y)$, and sf $S_{XY}(x, y)$. For the sake of simplicity, but without any loss of generality, we will always assume that the bivariate continuous rv is defined over the support $\mathbb{R}^+ \times \mathbb{R}^+$, i.e., outside this region $f(w, z)$ is zero; we also assume that $f(w, z)$ is finite.

3.1 Method I: Bivariate discretization preserving the joint pdf

Definition 1 (D-type bivariate discrete distribution) Given a bivariate continuous rv (W, Z) with pdf $f_{WZ}(w, z)$, its D-type bivariate discrete analogue is the rv (X, Y) with pmf

$$p(x, y) = f(x, y)/\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} f(i, j), \quad x \in \mathbb{N}, y \in \mathbb{N}. \tag{3}$$

It is trivial to show that (3) is a correctly defined pmf. The discretization method underlying Def.1 is the bivariate analogue of the univariate method recalled in Section 2.1: it preserves, up to a multiplicative factor, the expression of the joint pdf of the underlying continuous bivariate rv at any element of the support of its discrete analogue. The distribution generated using this technique may not always have a compact closed form due to the normalizing constant at the denominator of (3). The marginal pmf of X is given by

$$p_X(x) = \frac{\sum_j f(x, j)}{\sum_i \sum_j f(i, j)} = \frac{g(x)}{\sum_i g(i)},$$

with $g(i) = \sum_j f(i, j)$. Note that $p_X(x)$ in general is not the univariate discrete analogue of $f(x)$ according to the method outlined in Section 2.1; this would be

$$p_X^*(x) = \frac{f(x)}{\sum_{i=0}^{\infty} f(i)} = \frac{\int_0^{\infty} f(x, z) \mathrm{d}z}{\sum_{i=0}^{\infty} \int_0^{\infty} f(x, z) \mathrm{d}z}.$$

Symmetrical arguments hold for Y.

An interesting property of this discretization technique is the following: if the continuous bivariate distribution has independent components, then the corresponding bivariate discrete distribution has independent components too. In fact, if this hypothesis holds, we can write $f(w, z)$ as the product $f_W(w) f_Z(z)$ of the two marginal pdfs, and then

$$p(x, y) = \frac{f_W(x)}{\sum_{i \in \mathbb{N}} f_W(i)} \cdot \frac{f_Z(y)}{\sum_{j \in \mathbb{N}} f_Z(j)},$$

where the two factors on the right side are the pmfs of the components X and Y of the discrete bivariate rv; these are actually the univariate discrete analogues (according to the first univariate method of Section 2.1) of the continuous univariate components W and Z.

3.2 Method II: Bivariate discretization preserving the joint sf

Definition 2 (S-type bivariate discrete distribution) Given a bivariate continuous rv (W, Z) with sf $S_{W,Z}(w, z)$, its S-type bivariate discrete analogue

is the rv (X, Y) with pmf

$$p(x, y) = S_{WZ}(x, y) + S_{WZ}(x+1, y+1) - S_{WZ}(x, y+1) - S_{WZ}(x+1, y)$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{1} (-1)^{i+j} S(x+i, y+j), \quad x, y \in \mathbb{N}. \tag{4}$$

It can be easily proved that the function (4) is a valid pmf, i.e., $p(x, y) \geq 0$ $\forall (x, y) \in \mathbb{N} \times \mathbb{N}$ and $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) = 1$.

Proof The former property is trivial to prove, recalling and reformulating the rectangle inequality for bivariate rvs: the right member of (4) is in fact equal to $F_{WZ}(x, y) + F_{WZ}(x+1, y+1) - F_{WZ}(x, y+1) - F_{WZ}(x+1, y) = P(x < W < x+1, y < Z < y+1) \geq 0 \ \forall (x, y) \in \mathbb{R}^+ \times \mathbb{R}^+$.

As to the latter, exploiting telescoping properties, we have

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} S_{WZ}(i, j) - S_{WZ}(i, j+1) + S_{WZ}(i+1, j+1) - S_{WZ}(i+1, j)$$

$$= \sum_{i=0}^{\infty} S_{WZ}(i, 0) - S_{WZ}(i+1, 0) = S_{WZ}(0, 0) = P(W \geq 0, Z \geq 0) = 1.$$

The discretization method underlying Def.2 is the bivariate analogue of the univariate method recalled in Section 2.2. In fact, it can be proved the following result.

Proposition 1 *The S-type bivariate discrete rv (X, Y) preserves the sf of the parent bivariate continuous rv (W, Z) for any pair of non-negative integers, i.e., $S_{XY}(h, k) = S_{WZ}(h, k)$ for any $(h, k) \in \mathbb{N} \times \mathbb{N}$.*
The S-type distribution also preserves the expression of the sfs of the marginal components: $S_X(h) = S_W(h), \forall h \in \mathbb{N}$; $S_Y(k) = S_Z(k), \forall k \in \mathbb{N}$.

Proof For any non-negative integers h and k we have:

$$S_{XY}(h, k) = \sum_{i=h}^{\infty} \sum_{j=k}^{\infty} p(i, j)$$

$$= \sum_{i=h}^{\infty} \sum_{j=k}^{\infty} S_{WZ}(i, j) - S_{WZ}(i, j+1) + S_{WZ}(i+1, j+1) - S_{WZ}(i+1, j)$$

$$= \sum_{i=h}^{\infty} S_{WZ}(i, k) - S_{WZ}(i+1, k) = S_{WZ}(h, k).$$

Then, we also obtain, $\forall h, k \in \mathbb{N}$,

$$S_X(h) = S_{XY}(h, 0) = S_{WZ}(h, 0) = S_W(h) \tag{5a}$$

$$S_Y(k) = S_{XY}(0, k) = S_{WZ}(0, k) = S_Z(k). \tag{5b}$$

Exploiting the results of Proposition 1, we can state the following:

Proposition 2 *The marginal components X and Y of the S-type bivariate discrete rv (X, Y) are the discrete analogues (according to the method of Section 2.2) of the marginal components W and Z of the parent bivariate continuous rv (W, Z).*

Proof By using (5a), we have that

$$p_X(x) = S_X(x) - S_X(x+1) = S_W(x) - S_W(x+1) \qquad (6)$$

and

$$F_X(x) = 1 - P(X \geq x+1) = 1 - S_X(x+1) = 1 - S_W(x+1) = F_W(x+1);$$

and similarly for Y. (6) means that the marginal discrete distribution of X is actually the discrete analogue of the marginal continuous distribution of W according to the univariate method of Section 2.2 synthesized by (2). Analogous argument holds for Y.

From the two results above, another property descends.

Proposition 3 *If the bivariate continuous rv (W, Z) has copula $\mathcal{C}$, then its S-type bivariate discrete analogue (X, Y) has the same copula $\mathcal{C}$ on $\mathrm{Ran}(F_X) \times \mathrm{Ran}(F_Y)$, with $\mathrm{Ran}(F) := F(\bar{\mathbb{R}})$ being the range of F and $\bar{\mathbb{R}} = [-\infty, +\infty]$.*

First recall that we can write

$$F_{WZ}(w, z) = \mathcal{C}(F_W(w), F_Z(z))$$

for any $(w, z) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}}$ for some copula $\mathcal{C}$ (Sklar, 1959). This copula is unique, being (W, Z) a continuous random vector. Then

$$
\begin{aligned}
F_{XY}(h, k) &= P(X \leq h) + P(Y \leq k) - 1 + P(X \geq h+1, Y \geq k+1) \\
&= 1 - S_X(h+1) + 1 - S_Y(k+1) - 1 + S_{XY}(h+1, k+1) = \\
&= 1 - S_W(h+1) + 1 - S_Z(k+1) - 1 + S_{WZ}(h+1, k+1) = \\
&= F_{WZ}(h+1, k+1) = \mathcal{C}(F_W(h+1), F_Z(k+1)) = \mathcal{C}(F_X(h), F_Y(k))
\end{aligned}
$$

$\forall (h, k) \in \mathbb{N} \times \mathbb{N}$; and this means that (X, Y) has the same copula as (W, Z) for all the points $\mathrm{Ran}(F_X) \times \mathrm{Ran}(F_Y) \subset [0, 1]^2$; please however note that $\mathcal{C}$ is not the unique possible copula for (X, Y) over $[0, 1]^2$ (Sklar 1959, McNeil et al. 2005, pp.186-188). Thus, the S-type bivariate discrete analogue, by preserving the joint sf of its continuous parent distribution, also preserves its dependence structure, represented by the copula $\mathcal{C}$.

We can also state, in analogy with the univariate case, that the distribution of the S-type discrete analogue of (W, Z) coincides with that of $(\lfloor W \rfloor, \lfloor Z \rfloor)$.

Proposition 4 *The S-type bivariate discrete rv (X, Y) corresponding to the bivariate continuous rv (W, Z) has the same distribution as $(\lfloor W \rfloor, \lfloor Z \rfloor)$.*

Proof $(\lfloor W \rfloor, \lfloor Z \rfloor)$ clearly takes values over $\mathbb{N} \times \mathbb{N}$ and then, recalling Proposition 1,

$$
\begin{aligned}
P(\lfloor W \rfloor = h, \lfloor Z \rfloor = k) &= P(h \leq W < h+1, k \leq Z < k+1) \\
&= S_{WZ}(h,k) + S_{WZ}(h+1, h+1) - S_{WZ}(h, k+1) - S_{WZ}(h+1, k) \\
&= p_{XY}(h,k)
\end{aligned}
$$

for any pair (h, k) of non-negative integers.

This result appears consistent with the findings of Proposition 3: since the floor function is a monotone (non-strictly) increasing function, and since copulas are invariant to (strictly) increasing transformations, it is predictable that (X, Y) – or $(\lfloor W \rfloor, \lfloor Z \rfloor)$ – has the same copula as (W, Z), at least on a subset of $[0, 1]^2$.

   The discretization technique yielding the S-type discrete analogue shares a property with the method of Section 3.1: if the bivariate continuous distribution has independent components, then the corresponding bivariate discrete distribution has independent components too. In fact, if the hypothesis holds, we can write $S_{wz}(w, z)$ as the product $S_W(w)S_Z(z)$ of the two marginal sfs, and then, from (4),

$$
\begin{aligned}
p(x, y) &= S_W(x)S_Z(y) + S_W(x+1)S_Z(y+1) - S_W(x)S_Z(y+1) - S_W(x+1)S_Z(y) \\
&= [S_W(x) - S_W(x+1)][S_Z(y) - S_Z(y+1)] = p_X(x) \cdot p_Y(y) \quad \forall (x, y) \in \mathbb{N} \times \mathbb{N}.
\end{aligned}
$$

4 Applications

In this section, we consider two bivariate continuous distributions and derive their discrete analogues according to the two techniques presented in Sections 3.1, Def.1, and Section 3.2, Def.2, highlighting the theoretical aspects there described and underlining their peculiarities and differences.

4.1 Gumbel's type I bivariate exponential distribution

We consider the bivariate exponential distribution with joint pdf

$$
f(w, z) = \lambda_1 \lambda_2 [(1+\lambda\lambda_1 w)(1+\lambda\lambda_2 z) - \lambda] e^{-(\lambda_1 w + \lambda_2 z + \lambda\lambda_1\lambda_2 wz)} \quad w, z \in \mathbb{R}^+, \; (7)
$$

with $\lambda_1, \lambda_2 > 0$ and $0 \leq \lambda \leq 1$, usually referred to as Gumbel (type I) bivariate exponential distribution (see Gumbel (1960); Barnett (1980); Rodrigues et al (2011); and Balakrishnan and Lai (2009, pp.92-95)), whose joint cdf is given by:

$$
F_{WZ}(w, z) = 1 - e^{-\lambda_1 w} - e^{-\lambda_2 z} + e^{-(\lambda_1 w + \lambda_2 z + \lambda\lambda_1\lambda_2 wz)}
$$

and whose joint sf has the following expression:

$$
S_{WZ}(w, z) = e^{-\lambda_1 w - \lambda_2 z - \lambda\lambda_1 w \lambda_2 z}, \tag{8}
$$

to which corresponds the Gumbel-Barnett survival copula (Balakrishnan and Lai, 2009, p.95) $\hat{\mathcal{C}}(u, v) = uve^{-\lambda \ln u \ln v}$. It is easy to show that the marginal distributions of W and Z are exponential with parameter λ_1 and λ_2, respectively. The linear correlation between W and Z is given by

$$\rho_{WZ} = -1 + \int_0^{+\infty} \frac{e^{-x}}{1 + \lambda x} dx, \tag{9}$$

so it does not depend on either λ_1 or λ_2 and is a decreasing function of λ, taking the maximum value 0 if $\lambda = 0$ and the minimum value -0.40365 if $\lambda = 1$.

Applying the first discretization methodology to the joint pdf (7), we have that the normalizing constant is given by

$$\begin{aligned}
C &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \lambda_1 \lambda_2 [1 - \lambda + \lambda\lambda_1 i + \lambda\lambda_2 j + \lambda^2 \lambda_1 \lambda_2 ij] e^{-\lambda_1 i} (e^{-\lambda_2 - \lambda\lambda_1\lambda_2 i})^j \\
&= (1-\lambda)\lambda_1\lambda_2 \sum_{i=0}^{\infty} \frac{e^{-\lambda_1 i}}{1 - e^{-\lambda_2 - \lambda\lambda_1\lambda_2 i}} + \lambda^2 \lambda_1^2 \lambda_2^2 e^{-\lambda_2} \sum_{i=0}^{\infty} \frac{ie^{-\lambda_1 i}e^{-\lambda\lambda_1\lambda_2 i}}{(1 - e^{-\lambda_2 - \lambda\lambda_1\lambda_2 i})^2} \\
&+ \lambda\lambda_1^2 \lambda_2 \sum_{i=0}^{\infty} \frac{ie^{-\lambda_1 i}}{1 - e^{-\lambda_2 - \lambda\lambda_1\lambda_2 i}} + \lambda\lambda_2^2 \lambda_1 e^{-\lambda_2} \sum_{i=0}^{\infty} \frac{e^{-\lambda_1 i}e^{-\lambda\lambda_1\lambda_2 i}}{(1 - e^{-\lambda_2 - \lambda\lambda_1\lambda_2 i})^2}, \tag{10}
\end{aligned}$$

which cannot be reduced to an analytical closed form. The joint pmf of the D-type bivariate discrete rv can be then written as

$$p(x, y) = C^{-1} \left\{ \ln\theta_1 \ln\theta_2 \theta_1^x \theta_2^y \theta_3^{xy} [(1 + x\ln\theta_3/\ln\theta_2)(1 + y\ln\theta_3/\ln\theta_1) + \ln\theta_3/(\ln\theta_1\ln\theta_2)] \right\}, \tag{11}$$

where $\theta_1 = e^{-\lambda_1}$, $\theta_2 = e^{-\lambda_2}$, $\theta_3 = e^{-\lambda\lambda_1\lambda_2}$, or also as

$$p(x, y) = C^{*-1} \left\{ (1-\theta_1)(1-\theta_2)\theta_1^x \theta_2^y \theta_3^{xy} [(1 + x\ln\theta_3/\ln\theta_2)(1 + y\ln\theta_3/\ln\theta_1) + \ln\theta_3/(\ln\theta_1\ln\theta_2)] \right\}, \tag{12}$$

being $C^* = C[(1-\theta_1)(1-\theta_2)]/(\ln\theta_1\ln\theta_2)$. Note that the bivariate discrete distribution defined by (11) or (12) reduces to the product of two geometric pmfs with parameters θ_1 and θ_2 when $\theta_3 = 1$ (corresponding to $\lambda = 0$, i.e., independent components for the bivariate exponential distribution).

The marginal pmf of X can be derived by marginalizing (11) and is given by

$$p_X(x) = C^{-1}\theta_1^x \ln\theta_1 \ln\theta_2 \cdot$$
$$\left[\left(1 + \frac{\ln\theta_3}{\ln\theta_1\ln\theta_2}\right) \frac{1}{1 - \theta_2\theta_3^x} + \frac{\ln\theta_3^2}{\ln\theta_1\ln\theta_2} \frac{x\theta_2\theta_3^x}{(1 - \theta_2\theta_3^x)^2} + \frac{\ln\theta_3}{\ln\theta_2} \frac{x}{1 - \theta_2\theta_3^x} + \frac{\ln\theta_3}{\ln\theta_1} \frac{\theta_2\theta_3^x}{(1 - \theta_2\theta_3^x)^2} \right],$$

and is geometric with parameter θ_1 if and only if $\theta_3 = 1$.

Applying the second discretization methodology to the sf in (8), we have

$$p(x, y) = e^{-\lambda_1 x - \lambda_2 y - \lambda \lambda_1 x \lambda_2 y} + e^{-\lambda_1(x+1) - \lambda_2(y+1) - \lambda \lambda_1(x+1)\lambda_2(y+1)}$$

$$- e^{-\lambda_1 x - \lambda_2(y+1) - \lambda \lambda_1 x \lambda_2(y+1)} - e^{-\lambda_1(x+1) - \lambda_2 y - \lambda \lambda_1(x+1)\lambda_2 y}$$

$$= \theta_1^x \theta_2^y \theta_3^{xy}[1 + \theta_1 \theta_2 \theta_3 \theta_3^{x+y} - \theta_2 \theta_3^x - \theta_1 \theta_3^y], \qquad (13)$$

again with $\theta_1 = e^{-\lambda_1}$, $\theta_2 = e^{-\lambda_2}$, $\theta_3 = e^{-\lambda \lambda_1 \lambda_2}$. This pmf corresponds to that of the bivariate geometric distribution introduced by Roy (1993), which is the only bivariate discrete rv possessing locally constant bivariate failure rates. We remark that this distribution actually has geometric margins with parameter θ_1 and θ_2; and that the geometric distribution is the discrete counterpart (according to both discretization methods of Sections 2.1 and 2.2) of the exponential distribution. If $\lambda = 0$ ($\theta_3 = 1$), i.e., if the two components of the bivariate continuous model are independent, the above joint pmf describes a bivariate geometric rv with independent components. Note that in this case, the two discretization methodologies of Sections 3.1 and 3.2 yield - as one could have expected - the same bivariate discrete distribution.

Tables 1a and 1b partially report the joint pmfs for the D-type and S-type analogues derived from Gumbel's type I bivariate exponential distribution with parameters $\lambda_1 = 1$, $\lambda_2 = 1/3$, and $\lambda = 1/2$. For this choice of parameters of the parent distribution, the differences in the joint probabilities of the two discrete analogues are not negligible; moreover, the marginal distributions of X and Y for the D-type analogue are quite different from their counterparts for the S-type (which we know are geometric). The linear correlation for the continuous model, recalling (9), is equal to -0.2773; for the D-type analogue it can be computed numerically and is equal to -0.2991; for the S-type, -0.2471.

4.2 FGM bivariate exponential distribution

Consider the bivariate cdf:

$$F_{WZ}(w, z) = (1 - e^{-\lambda_1 w})(1 - e^{-\lambda_2 z})(1 + \theta e^{-\lambda_1 w - \lambda_2 z}) \quad w, z \in \mathbb{R}^+ \qquad (14)$$

whose marginal distributions W and Z are exponential with parameter λ_1 and λ_2, respectively, and whose copula is the Farlie-Gumbel-Morgenstern (FGM) copula (Farlie, 1960) with parameter $\theta \in [-1, +1]$:

$$\mathcal{C}(u, v) = uv[1 + \theta(1 - u)(1 - v)].$$

This distribution is referred to as Gumbel's type II bivariate exponential distribution or simply FGM bivariate exponential distribution. Letting $\theta = 0$ yields two independent components; positive (negative) values of θ yield positively (negatively) correlated components. The correlation coefficient is in fact give by $\rho_{WZ} = \theta/4$, so it does not depend on the marginal parameters λ_1 and λ_2,

Table 1: Joint pmfs of the D- and S-type discrete analogues of Gumbel's type I bivariate exponential distribution (with $\lambda_1 = 1$, $\lambda_2 = 1/3$, $\lambda = 1/2$)

(a) D-type

x, y	0	1	2	3	$\cdots$	$P(X = x)$
0	0.0926	0.0885	0.0793	0.0681	$\cdots$	0.6021
1	0.0681	0.0517	0.0376	0.0266	$\cdots$	0.2399
2	0.0376	0.0236	0.0143	0.0085	$\cdots$	0.0954
3	0.0184	0.0097	0.0049	0.0025	$\cdots$	0.0378
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$P(Y = y)$	0.2318	0.1794	0.1384	0.1066	$\cdots$	1

(b) S-type

x, y	0	1	2	3	$\cdots$	$P(X = x)$
0	0.1387	0.1153	0.0923	0.0720	$\cdots$	0.6321
1	0.0789	0.0540	0.0359	0.0234	$\cdots$	0.2325
2	0.0377	0.0216	0.0120	0.0066	$\cdots$	0.0855
3	0.0166	0.0080	0.0037	0.0017	$\cdots$	0.0315
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$P(Y = y)$	0.2835	0.2031	0.1455	0.1043	$\cdots$	1

and due to the bounds for θ we have that $-1/4 \leq \rho_{WZ} \leq 1/4$. The joint pdf corresponding to (14) can be easily derived and is equal to:

$$f(w, z) = \lambda_1 \lambda_2 e^{-\lambda_1 w - \lambda_2 z}[1 + \theta(1 - 2e^{-\lambda_1 w})(1 - 2e^{-\lambda_2 z})], \qquad (15)$$

whereas the joint sf has the following expression:

$$S_{WZ}(w, z) = e^{-\lambda_1 w - \lambda_2 z}[1 + \theta(1 - e^{-\lambda_1 w})(1 - e^{-\lambda_2 z})]. \qquad (16)$$

Applying the first discretization methodology to (15), we have that the normalizing constant is given by:

$$C = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \lambda_1 \lambda_2 e^{-\lambda_1 i - \lambda_2 j}[1 + \theta(1 - 2e^{-\lambda_1 i})(1 - 2e^{-\lambda_2 j})] \qquad (17)$$

$$= \frac{(1 + \theta)\lambda_1 \lambda_2}{(1 - e^{-\lambda_1})(1 - e^{-\lambda_2})} + \frac{4\theta \lambda_1 \lambda_2}{(1 - e^{-2\lambda_1})(1 - e^{-2\lambda_2})}$$

$$- \frac{2\theta \lambda_1 \lambda_2}{(1 - e^{-\lambda_1})(1 - e^{-2\lambda_2})} - \frac{2\theta \lambda_1 \lambda_2}{(1 - e^{-2\lambda_1})(1 - e^{-\lambda_2})}.$$
$$(18)$$

Letting $\theta_1 = e^{-\lambda_1}$ and $\theta_2 = e^{-\lambda_2}$, we can rewrite C as

$$C = \ln \theta_1 \ln \theta_2 \left[\frac{1 + \theta}{(1 - \theta_1)(1 - \theta_2)} + \frac{4\theta}{(1 - \theta_1^2)(1 - \theta_2^2)} - \frac{2\theta}{(1 - \theta_1)(1 - \theta_2^2)} - \frac{2\theta}{(1 - \theta_1^2)(1 - \theta_2)} \right]$$

$$= \ln \theta_1 \ln \theta_2 \frac{\theta(1 - \theta_1)(1 - \theta_2) + (1 + \theta_1)(1 + \theta_2)}{(1 - \theta_1^2)(1 - \theta_2^2)} = C^* \ln \theta_1 \ln \theta_2 \qquad (19)$$

with $C^* = C/(\ln\theta_1 \ln\theta_2)$. Thus the joint pmf of the D-type discrete rv (X, Y) can be written as

$$p(x, y) = (\ln\theta_1 \ln\theta_2)\theta_1^x\theta_2^y[1 + \theta(1 - 2\theta_1^x)(1 - 2\theta_2^y)]/C = \theta_1^x\theta_2^y[1 + \theta(1 - 2\theta_1^x)(1 - 2\theta_2^y)]/C^*. \tag{20}$$

The marginal pmf of X can be derived as:

$$p_X(x) = \sum_{y=0}^{\infty} p(x, y) = \left[\frac{\theta_1^x}{1 - \theta_2} + \theta\theta_1^x(1 - 2\theta_1^x)\left(\frac{1}{1 - \theta_2} - \frac{2}{1 - \theta_2^2}\right)\right]/C^*$$

$$= \left[\theta_1^x\left(\frac{1}{1 - \theta_2} - \frac{\theta}{1 + \theta_2}\right) + \theta_1^{2x}\frac{2\theta}{1 + \theta_2}\right]/C^*$$

and is thus geometric if and only if $\theta = 0$ (independence case). If θ is larger than zero, than the marginal distribution of X is a mixture of two geometric distributions G_1 and G_2, with parameters θ_1 and θ_1^2, and weights $w_1 = (1 + \theta_2 - \theta + \theta\theta_2)/[(1 - \theta_2)(1 + \theta_2)(1 - \theta_1)]$ and $w_2 = 2\theta/[(1 + \theta_2)(1 - \theta_1)(1 + \theta_1)]$, respectively. Note that the marginal pmf of X depends not only on the parameter λ_1 (θ_1) of the marginal distribution of W, but also on the other marginal parameter λ_2 (θ_2) and on the dependence parameter θ.

We can easily derive the joint cdf of (X, Y):

$$F_{XY}(x, y) = \sum_{j=0}^{y}\sum_{i=0}^{x} 1/C^*[\theta_1^i\theta_2^j + \theta\theta_1^i(1 - 2\theta_1^i)\theta_2^j(1 - 2\theta_2^j)]$$

$$= 1/C^* \sum_{j=0}^{y} \frac{1 - \theta_1^{x+1}}{1 - \theta_1}\theta_2^j + \theta\theta_2^j(1 - 2\theta_2^j)[\frac{1 - \theta_1^{x+1}}{1 - \theta_1} - 2\frac{1 - \theta_1^{2(x+1)}}{1 - \theta_1^2}]$$

$$= 1/C^* \left\{\frac{1 - \theta_1^{x+1}}{1 - \theta_1}\frac{1 - \theta_2^{y+1}}{1 - \theta_2} + \theta\left[\frac{1 - \theta_2^{y+1}}{1 - \theta_2} - 2\frac{1 - \theta_2^{2(y+1)}}{(1 - \theta_2^2)}\right]\left[\frac{1 - \theta_1^{x+1}}{1 - \theta_1} - 2\frac{1 - \theta_1^{2(x+1)}}{(1 - \theta_1^2)}\right]\right\}$$

$$= 1/C^* \cdot \frac{1 - \theta_1^{x+1}}{1 - \theta_1} \cdot \frac{1 - \theta_2^{y+1}}{1 - \theta_2}\left[1 + \theta\left(1 - \frac{2(1 + \theta_2^{y+1})}{1 + \theta_2}\right)\left(1 - \frac{2(1 + \theta_1^{x+1})}{1 + \theta_1}\right)\right]$$

and the marginal cdf of X, which is is equal to:

$$F_X(x) = 1/C^* \cdot \frac{1 - \theta_1^{x+1}}{1 - \theta_1}\frac{1}{1 - \theta_2}\left[1 + \theta\frac{\theta_2 - 1}{1 + \theta_2}\left(1 - \frac{2(1 + \theta_1^{x+1})}{1 + \theta_1}\right)\right]$$

$$= \left[\frac{1 - \theta_1^{x+1}}{1 - \theta_1}\left(\frac{1}{1 - \theta_2} - \frac{\theta}{1 + \theta_2}\right) + \frac{1 - \theta_1^{2(x+1)}}{1 - \theta_1^2}\frac{2\theta}{1 + \theta_2}\right]/C^*.$$

For a given $0 < u < 1$, we can derive the u-quantile of X, by equating $F_X(x)$ to u and solving this equation with respect to x, thus obtaining a root, say x_u^*, and then taking the smallest integer larger than x_u^*, $\lceil x_u^* \rceil$. The u-quantile is then given by $x_u = \lceil \ln z/\ln\theta_1 \rceil - 1$, with

$$z = \frac{-b + \sqrt{(b + 2a)^2 - 4ac}}{2a}$$

where $a = \dfrac{2\theta}{(1 - \theta_1^2)(1 + \theta_2)}$, $b = \dfrac{1}{1 - \theta_1}\left(\dfrac{1}{1 - \theta_2} - \dfrac{\theta}{1 + \theta_2}\right)$, $c = C^*u$.

We also have

$$\mathbb{E}(X) = \frac{1}{C^*}\left[\frac{\theta_1}{(1 - \theta_1)^2}\left(\frac{1}{1 - \theta_2} - \frac{\theta}{1 + \theta_2}\right) + \frac{\theta_1^2}{(1 - \theta_1^2)^2}\frac{2\theta}{1 + \theta_2}\right];$$

$$\mathbb{E}(XY) = \frac{1}{C^*}\left[\frac{\theta_1}{(1 - \theta_1)^2}\frac{\theta_2}{(1 - \theta_2)^2} + \theta\left(\frac{\theta_1}{(1 - \theta_1)^2} - \frac{2\theta_1^2}{(1 - \theta_1^2)^2}\right)\left(\frac{\theta_2}{(1 - \theta_2)^2} - \frac{2\theta_2^2}{(1 - \theta_2^2)^2}\right)\right];$$

$$\mathbb{E}(X^2) = \frac{1}{C^*}\left[\frac{\theta_1(1 + \theta_1)}{(1 - \theta_1)^3}\left(\frac{1}{1 - \theta_2} - \frac{\theta}{1 + \theta_2}\right) + \frac{2\theta\theta_1^2(1 + \theta_1^2)}{(1 - \theta_1^2)^3(1 + \theta_2)}\right];$$

analogous results hold for Y. The expression of the linear correlation can be then easily derived; here we report it when X and Y are identically distributed, i.e., when $\theta_1 = \theta_2$:

$$\rho_{XY} = \frac{4\theta\theta_1(1 + \theta_1)^2}{[(1 + \theta_1)^6 - 2\theta\theta_1(1 - \theta_1)\,(\theta_1^2 + 2\theta_1 - 1)\,(1 + \theta_1)^2 + \theta^2(1 - \theta_1)^2\,(\theta_1^4 + 2\theta_1^3 - 4\theta_1^2 - 2\theta_1 - 1)]}. \tag{21}$$

It can be proved that ρ_{XY}, for θ_1 fixed in $(0, 1)$, is an increasing function of θ in $[-1, 1]$. In fact, if we compute the first order partial derivative of ρ_{XY} with respect to θ, we obtain:

$$\frac{\partial \rho_{XY}}{\partial \theta} = 4\theta_1(1 + \theta_1)^2 \cdot \frac{g(\theta_1, \theta) - \theta g_\theta'(\theta_1, \theta)}{g(\theta_1, \theta)^2},$$

with $g(\theta_1, \theta)$ being the denominator of (21) and $g_\theta'(\theta_1, \theta)$ the partial derivative of $g(\theta_1, \theta)$ with respect to θ. In order to study the sign of $\partial \rho_{XY}/\partial \theta$ it is sufficient to study the sign of $g(\theta_1, \theta) - \theta g_\theta'(\theta_1, \theta)$, for which we have

$$g(\theta_1, \theta) - \theta g_\theta'(\theta_1, \theta) = \theta_1^6(1 - \theta^2) + 6\theta_1^5 + \theta_1^4(15 + 7\theta^2) + \theta_1^2(20 - 8\theta^2) + \theta_1^2(15 + \theta^2) + 1 + \theta^2,$$

which is clearly always positive for any choice of $\theta_1 \in (0, 1)$ and $\theta \in [-1, 1]$; so ρ_{XY} is an increasing function of θ for any fixed value of θ_1. Thus, although entering the expressions of the marginal pmfs, θ is responsible for the direction and strength of the (linear) correlation between X and Y, once θ_1 and θ_2 are held fixed.

In Figure 1, the minimum and maximum attainable correlations are plotted for the D-type discrete version of the FGM exponential distribution with identically distributed margins ($\theta_1 = \theta_2$), by setting the value of the dependence parameter θ equal to -1 and $+1$, respectively. It can be proved that the maximum value of ρ_{XY} is approximately equal to 0.5149 and is reached at $\theta = 1$ and $\theta_1 \approx 0.1251$; the infimum value is -1 and is attained when $\theta = -1$ and $\theta_1 = \theta_2 \to 0^+$; in this case, (X, Y) tends to a bivariate discrete distribution with possible values $(0, 1)$ and $(1, 0)$ and associated probability $1/2$ each; the margins are both Bernoulli with parameter $p = 0.5$ and the correlation is -1.

Applying the second discretization methodology to (16) we obtain the following joint pmf for the discrete analogue (X, Y):

$$
\begin{aligned}
p(x, y) = {} & e^{-\lambda_1 x - \lambda_2 y}[1 + \theta(1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2 y})] \\
& + e^{-\lambda_1(x+1) - \lambda_2(y+1)}[1 + \theta(1 - e^{-\lambda_1(x+1)})(1 - e^{-\lambda_2(y+1)})] \\
& - e^{-\lambda_1 x - \lambda_2(y+1)}[1 + \theta(1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2(y+1)})] \\
& - e^{-\lambda_1(x+1) - \lambda_2 y}[1 + \theta(1 - e^{-\lambda_1(x+1)})(1 - e^{-\lambda_2 y})],
\end{aligned}
$$

and then, substituting $\theta_1 = e^{-\lambda_1}$, $\theta_2 = e^{-\lambda_2}$, and after some algebraic steps, we obtain:

$$
p(x, y) = \theta_1^x(1 - \theta_1)\theta_2^y(1 - \theta_2)\left\{1 + \theta[1 - \theta_1^x(1 + \theta_1)][1 - \theta_2^y(1 + \theta_2)]\right\}. \quad (22)
$$

The above joint pmf corresponds to that of the bivariate geometric distribution described in Barbiero (2017, 2018a), after substituting θ_i with $1 - \theta_i$, $i = 1, 2$; we recall that it actually corresponds to the joint pmf of two geometric margins (with parameters θ_1 and θ_2) linked together through the FGM copula with parameter θ, i.e., the same copula of (W, Z), confirming the remark we made in Section 2.2. The correlation between X and Y is then given by

$$
\rho_{XY} = \theta \frac{\sqrt{\theta_1 \theta_2}}{(1 + \theta_1)(1 + \theta_2)},
$$

and, differently from ρ_{WZ}, its value depends not only on θ, but also on the marginal parameters θ_1 and θ_2; once the latter are fixed, ρ_{XY} is directly proportional to θ. Note that for this discrete distribution, the parameter θ can range over a larger interval than that allowed for the parent continuous distribution, namely $[-1, \min(1/\theta_1, 1/\theta_2)]$ (Piperigou, 2009). In Figure 2, the minimum and maximum attainable correlations are plotted for the S-type discrete version of the FGM exponential distribution with identically distributed margins ($\theta_1 = \theta_2$), by setting the value of the dependence parameter θ equal to -1 and $+1/\theta_1$.

Please note the similarity between the pmfs (20) and (22): they are both of the form $k \cdot \theta_1^x \theta_2^y[1 + \theta g_1(x, \theta_1)g_2(y, \theta_2)]$, with k being a real constant and g_1 and g_2 functions depending only on x and θ_1, y and θ_2, respectively. However, only if $\theta = 0$, i.e., if the continuous model (14) have independent exponentially distributed components, then also the discrete analogues (20) and (22) (which now coincide) have independent (geometric) components.

Just to have an idea of the difference between the discrete analogues (20) and (22) derived from the same continuous distribution, let us consider the distribution (14) with parameters $\lambda_1 = 1$, $\lambda_2 = 1/3$, and $\theta = 1/2$ (and then $\theta_1 = 0.3679$, $\theta_2 = 0.7165$). For the pmf in (20), the value of the normalizing constant C of (19) is approximately equal to 1.931243. In Tables 2a and 2b, the joint pmfs of the two discrete analogues are (partially) reported (only the values $x, y = 0, 1, 2, 3$ are considered for the sake of brevity). One can compare the two tables cell by cell in order to ascertain differences between the two joint distributions; please note that significant deviations exist also comparing

Fig. 1: Correlation bounds for the D-type discrete version of the FGM bivariate exponential distribution with identically distributed margins (solid lines); dashed lines indicate the bounds for the parent distribution ($-1/4$ and $1/4$)
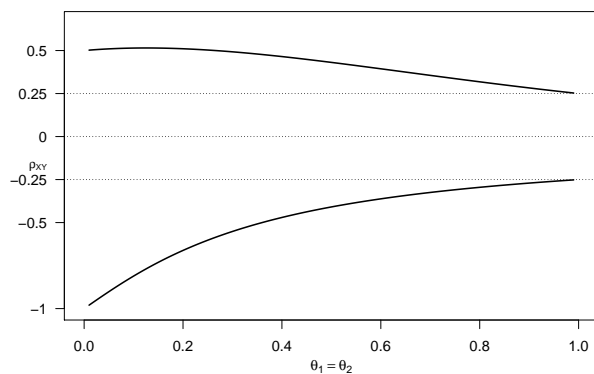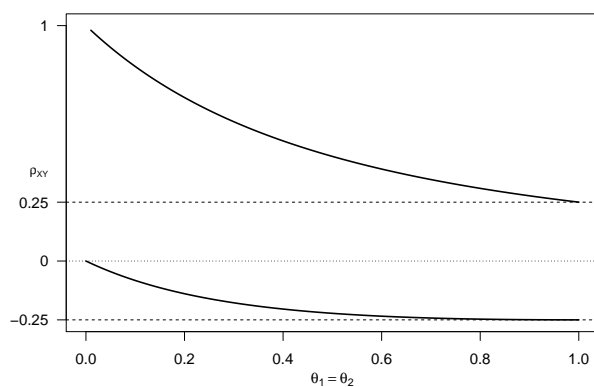

Fig. 2: Correlation bounds for the S-type discrete version of the FGM bivariate exponential distribution with identically distributed margins (solid lines); dashed lines indicate the bounds for the parent distribution ($-1/4$ and $1/4$). The θ parameter here ranges from -1 to $1/\theta_1$

homologous marginal distributions $p_X(x)$ and $p_Y(y)$ (remember that only in the second case the two margins are geometric).

5 A numerical example

In this section, we provide a practical simulation experiment, which refers to the Gumbel's type I bivariate exponential distribution discussed in Section 4.1. This simulation experiment illustrates how the bivariate discretization provided through the second method of Section 3.2 can be practically

Table 2: Joint pmfs of the D- and S-type discrete analogues of the FGM bivariate exponential distribution (with $\lambda_1 = 1$, $\lambda_2 = 1/3$, $\theta = 1/2$)

(a) D-type

x, y	0	1	2	3	$\cdots$	$P(X = x)$
0	0.2589	0.1505	0.0898	0.0551	$\cdots$	0.6592
1	0.0551	0.0429	0.0325	0.0242	$\cdots$	0.2191
2	0.0148	0.0141	0.0119	0.0094	$\cdots$	0.0774
3	0.0047	0.0050	0.0044	0.0035	$\cdots$	0.0281
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$P(Y = y)$	0.3361	0.2152	0.1411	0.0943	$\cdots$	1

(b) S-type

x, y	0	1	2	3	$\cdots$	$P(X = x)$
0	0.2028	0.1338	0.0900	0.0615	$\cdots$	0.6321
1	0.0542	0.0445	0.0348	0.0265	$\cdots$	0.2325
2	0.0172	0.0157	0.0131	0.0103	$\cdots$	0.0855
3	0.0059	0.0057	0.0048	0.0038	$\cdots$	0.0315
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$P(Y = y)$	0.2835	0.2031	0.1455	0.1043	$\cdots$	1

implemented. It is conducted under the R programming environment; relevant code is available as supplementary material.

We first draw a bivariate i.i.d. sample (w_i, z_i), $i = 1, \ldots, n$, of size $n = 100$ from the Gumbel's type I bivariate exponential distribution (W, Z), with the following choice of parameters: $\lambda_1 = 1/3$, $\lambda_2 = 1/4$, $\lambda = 1/2$. Based on this sample, we numerically compute the MLEs of the three parameters: $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}$. Then we discretize the original sample taking the integer part of each component, obtaining $(x_i = \lfloor w_i \rfloor, y_i = \lfloor z_i \rfloor)$, $i = 1, \ldots, n$. On this new series of paired values, which can be regarded as a sample from the bivariate geometric distribution by Roy (1993) with parameters $\theta_1 = e^{-\lambda_1}, \theta_2 = e^{-\lambda_2}, \theta = e^{-\lambda\lambda_1\lambda_2}$ (see again Section 4.1), we compute the MLEs $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}$. MLEs for the original and discretized sample distributions are reported in Table 3; Figure 3a displays the scatter-plot for the original sample; Figure 3b displays the bubble-plot for the "discretized" sample (in a bubble-plot, each "bubble" has an area proportional to the corresponding joint frequency of the bivariate observation).

The simulation procedure used for drawing a sample from the continuous bivariate distribution (7) is now illustrated. If we compute the conditional pdf of Z given $W = w$, we obtain

$$f_{Z|W}(z|w) = f_{WZ}(w, z)/f_W(w) = p\beta\lambda_2 e^{-\beta\lambda_2 z} + (1 - p)z\beta^2\lambda_2^2 e^{-\beta\lambda_2 z}$$

with $p = (\beta - \lambda)/\beta$ and $\beta = 1 + \lambda\lambda_1 w$, and we can easily recognize that the conditional pdf $f_{Z|W}$ is indeed a finite mixture between an exponential distribution with parameter $\beta\lambda_2$ and a Gamma distribution with parameters 2 and $\beta\lambda_2$, and weights p and $1-p$, respectively. Then, the following algorithm (which is a generalization of the algorithm described in Johnson 1987, pp.197-
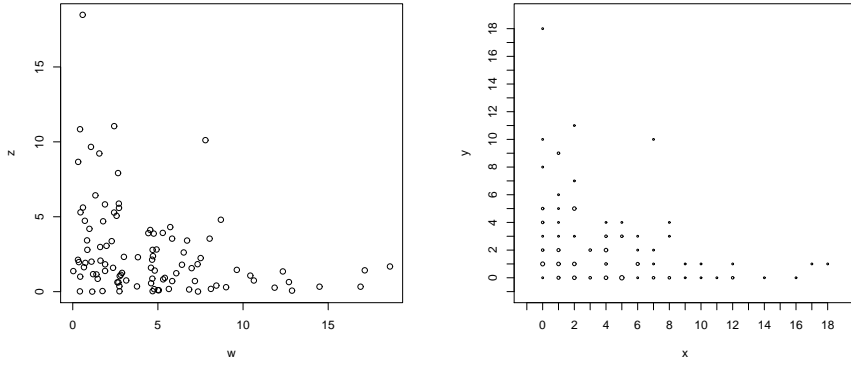
(a) Scatter plot of the original sample

(b) Bubble plot of the discretized sample: each "bubble" has an area proportional to the frequency of the corresponding pair (x, y)

Fig. 3: Graphical representation for the distribution of a random sample from the bivariate Gumbel's type I distribution with parameters $\lambda_1 = 1/3$, $\lambda_2 = 1/4$, and $\lambda = 1/2$ (a) and its S-type discretized version (b).

Table 3: Descriptive statistics and MLEs for the bivariate Gumbel exponential and Geometric distributions fitted to the bivariate samples displayed in Figures 3a and 3b

	continuous distribution			discretized (S-type) distribution		
parameter	λ_1	λ_2	λ	θ_1	θ_2	θ_3
true value	1/4	1/3	1/2	0.7788	0.7165	0.9592
MLE	0.2209	0.3621	0.6082	0.8009	0.6982	0.9462
component	W	Z		X	Y	
mean	4	3		3.521	2.528	
sd	4	3		3.990	2.986	
corr.	-0.2773			-0.2722		
sample mean	4.631	2.685		4.08	2.24	
sample sd	3.985	3.026		3.992	2.985	
sample corr.	-0.3153			-0.3178		

198, for the case of exponential margins of unitary mean) can be used to generate samples from a rv (W, Z) distributed as (7):

1. Generate U_1, U_2, U_3 independent uniform rvs in $(0, 1)$
2. Set $W = -\log U_1/\lambda_1$
3. Compute $\beta = 1 + \lambda\lambda_1 W$ and $p = (\beta - \lambda)/\lambda$
4. Set $Y = -\log U_2$
5. If $U_3 < p$ then set $Z = Y/(\beta\lambda_2)$,
   else generate U_4 uniform in $(0, 1)$ and set $Z = [Y - \log(U_4)]/(\beta\lambda_2)$

According to the algorithm, we first draw a realization w from the marginal distribution W, which is exponentially distributed with parameter λ_1, and then we draw a realization z from the conditional distribution of Z given $W = w$.

Simulating a random sample from Roy's bivariate geometric distribution can be thus carried out indirectly by first simulating a sample from Gumbel's bivariate exponential distribution (resorting to the algorithm above) and then discretizing it. This procedure will work for any consistent choice of the three parameters of the original distribution. However, Roy's distribution allows the dependence parameter θ_3 to range within a larger interval than that would be implicitly dictated by the range of λ; in fact, we have that θ_3 must satisfy the two constraints: $0 < \theta_3 \leq 1$ and $\theta_3 \geq (\theta_1 + \theta_2 - 1)/(\theta_1 \theta_2)$ (see Roy (1993) for details). For example, in our case, $\theta_3 = \mathrm{e}^{-\lambda \lambda_1 \lambda_2}$ would be bounded between $\mathrm{e}^{-\lambda_1 \lambda_2} = 0.9200$ (when $\lambda = 1$) and 1 ($\lambda = 0$); whereas for Roy's distribution the dependence parameter θ_3 falls between 0.8928 and 1. Thus, if we want to simulate samples from a discretized version of Gumbel's bivariate exponential distribution, which results into Roy's bivariate geometric distribution with $\mathrm{e}^{-\lambda_1 \lambda_2} \leq \theta_3 \leq 1$, we can resort to the procedure above (simulation of a "continuous" sample and then discretization); if we want to simulate samples from Roy's bivariate geometric for any possible consistent value of θ_3, we need an ad-hoc procedure, based again on a conditional approach, but now referred to the discrete random components (see Barbiero, 2018b, 2019).

This example shows how it is easy to obtain a sample from the S-type discrete analogue once a sample from the continuous distribution is available; on the contrary, this does not hold for the D-type discrete analogue, where a direct and distinct simulation procedure, possibly still based on conditional sampling, has to be implemented.

6 Data analysis

In this section, we fit bivariate discrete models obtained as discrete analogues from a bivariate continuous distribution to two datasets taken from the literature. The first dataset concerns correlated lifetimes, which are however measured on a discrete scale; the second one consists of authentic discrete data.

6.1 Discrete recurrence times

We analyze a dataset resulting from a study of 38 kidney dialysis patients. Originally considered in McGilchrist and Aisbett (1991), this study is concerned with the prevalence of infection at the catheter insertion point. Two recurrence times are measured for each patient. A catheter is inserted, and the first time to infection (X, in days) is measured. If the catheter is removed for reasons other than infection, then the first recurrence time is censored.

If infection occurs, the catheter is removed, the infection cleared, and then after some predetermined period the catheter is reinserted. The second time to infection, Y, is measured as time elapsed between the second insertion and the second infection or censoring. The second recurrence time is censored if either the catheter is removed for reasons other than infection or the follow-up period for the patient ends before infection occurs.

Several bivariate exponential and generalized exponential distributions have been applied in the literature to model these data (see, e.g., Mirhosseini et al., 2015). However, since both X and Y are recorded in days, a bivariate discrete model can be more appropriate. For our scope, we limit the attention only on the patients ($n = 23$) whose times to infection X and Y are both uncensored. The scatter-plot of these data is displayed in Figure 4. Summary statistics for this reduced dataset are:

$$\bar{x} = 107.39, \quad s_X = 139.22, \quad \bar{y} = 116.13, \quad s_Y = 138.92, \quad \hat{\rho} = 0.1910.$$

Fig. 4: Scatter-plot of times to first and second infection for the kidney data set in McGilchrist and Aisbett (1991)
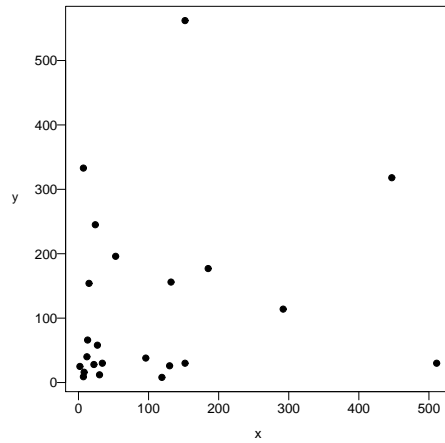
Due to the moderate level of correlation, we can first try and fit the FGM bivariate exponential distribution of Eq.(15) to the dataset; then, we can fit its two discrete analogues (20) and (22). Since the observed values of X and Y are quite large, we expect negligible differences in the goodness-of-fit moving from the continuous to the discrete models. We perform maximization of the log-likelihood function for the three models above and find the MLEs reported in Table 4, along with the maximum value of the log-likelihood functions ($\ell_{\max}$), and the Akaike Information Criterion (AIC), given by $2k - 2\ell_{\max}$, where k is the number of parameters (3 for all cases). It can be noted that the MLEs of the distribution parameters λ_1, λ_2 and θ are very close across the three models;

the original continuous model presents a slightly smaller value of the AIC, indicating a better fit. The explanation, as anticipated, is quite simple: in this case, where most observed values for both correlated variables are large, using a bivariate discrete model obtained as an analogue from a continuous one, does not introduce a considerable improvement in goodness-of-fit with respect to the original model. It is worth noting that the two discrete analogues themselves, fitted to these data, provide very close values for homologous parameters; indeed, large observed values for x and y are likely to produce estimates of θ_1 and θ_2 close to the boundary value 1, at least for the S-type distribution, whose margins are geometric; and for values of θ_1 and θ_2 close to 1, whatever the value of θ is, the expressions of the two pmfs (20) and (22) tend to coincide.

Table 4: MLEs, values of log-likelihood and AIC for the bivariate FGM bivariate exponential distribution and its discrete analogues applied to kidney data in McGilchrist and Aisbett (1991)

distribution, MLEs	$\hat{\lambda}_1$	$\hat{\theta}_1 = e^{-\hat{\lambda}_1}$	$\hat{\lambda}_2$	$\hat{\theta}_2 = e^{-\hat{\lambda}_2}$	$\hat{\theta}$	$\ell_{\max}$	AIC
Biv. FGM exp.	.00916	-	.00838	-	.5800	-262.17	530.34
D-type	.00912	.9909	.00834	.9917	.5823	-262.37	530.74
S-type	.00907	.9910	.00830	.9917	.5867	-262.38	530.76

6.2 Count data

The data, considered in Mitchell and Paulson (1981), consist of the number of aborts by 109 aircrafts in two (first $= x$, second $= y$) consecutive 6 months of 1-year period. Summary statistics for the dataset are $\bar{x} = 0.624$, $\bar{y} = 0.725$, $s_X^2 = 1.024$, $s_Y^2 = 1.062$. The sample correlation coefficient between x and y is $\hat{\rho}_{xy} = -0.1609$, which denotes a slight negative dependence.

The geometric distribution could be a plausible model for both margins x and y; since we know that the FGM copula can model slight negative or positive dependence, we can resort to the two discrete analogues of the FGM bivariate exponential rv, whose pmfs are expressed by Eqs.(20) and (22), and fit them to the dataset at study.

By using the maximum likelihood method, for both bivariate distributions we derive the parameters' estimates, along with the values of the maximized log-likelihood function $\ell_{\max}$ and AIC. These results are reported in Table 5 and indicate that the two models present a very close goodness-of-fit (the D-type model is preferable, but the two AIC are equal at the first decimal digit) although the parameter estimates (in particular for the θ parameter) and then the pmfs are quite different. We recall that while the S-type bivariate discrete distribution in this case assures that the univariate margins are geometrically distributed, this is not true for the D-type homologue. The theoretical frequencies for both bivariate discrete models, reconstructed by using the MLEs

Table 5: Results of the estimation for the FGM-type bivariate discrete distributions of Section 4.2 on the data taken from Mitchell and Paulson (1981)

distribution, estimates	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}$	$\ell_{\max}$	AIC
D-type	0.3566	0.3935	-0.2441	-244.6191	495.2382
S-type	0.3824	0.4250	-0.6174	-244.6304	495.2608

Table 6: Bivariate frequency distribution of the data taken from Mitchell and Paulson (1981): number of flight aborts by 109 aircrafts in the first and second consecutive six months of a one-year period. Observed frequencies are reported on the first row of each three-row group; theoretical joint frequencies fitted by the D-type and S-type discrete distributions derived from the FGM exponential distribution are reported on each second row between round brackets and third row between square brackets, respectively. The solid borders identify the cell groups that are used for the computation of the χ^2 statistic

$x \setminus y$	0	1	2	3	≥ 4	tot
0	34	20	4	6	4	68
	(33.86)	(18.54)	(8.10)	(3.31)	(2.19)	(66.01)
	[34.82]	[17.98]	[8.22]	[3.60]	[2.70]	[67.32]
1	17	7	0	0	0	24
	(17.09)	(6.19)	(2.35)	(0.91)	(0.59)	(27.14)
	[16.63]	[5.57]	[2.10]	[0.84]	[0.60]	[25.74]
2	6	4	1	0	0	11
	(6.73)	(2.15)	(0.77)	(0.29)	(0.19)	(10.13)
	[6.84]	[1.94]	[0.65]	[0.24]	[0.17]	[9.84]
3	0	4	0	0	0	4
	(2.48)	(0.76)	(0.27)	(0.10)	(0.06)	(3.67)
	[2.69]	[0.71]	[0.23]	[0.08]	[0.06]	[3.76]
4	0	0	0	0	0	0
	(0.90)	(0.27)	(0.09)	(0.03)	(0.02)	(1.32)
	[1.04]	[0.27]	[0.08]	[0.03]	[0.02]	[1.44]
≥ 5	2	0	0	0	0	2
	(0.50)	(0.15)	(0.05)	(0.02)	(0.01)	(0.73)
	[0.65]	[0.17]	[0.05]	[0.02]	[0.01]	[0.89]
tot	59	35	5	6	4	109
	(61.56)	(28.07)	(11.64)	(4.67)	(3.06)	(109)
	[62.67]	[26.64]	[11.32]	[4.81]	[3.56]	[109]

of Table 5, are displayed in Table 6 and compared with the corresponding observed values. At a glance, discrepancies are acceptable and the two models seem to be adequate. Absolute goodness-of-fit measures, such as the customary chi-square statistic, can be computed along with the associated p-value; however, the problem arises of how to aggregate cells in the theoretical contingency table in order to ensure a minimum value for each frequency (usually, a threshold of 5 is used). A possible aggregation of cells is highlighted in Table 6, which leads to $g = 8$ groupings. The chi-square statistic computed on this cell aggregation takes the value 6.8841 (p-value 0.1421) for the D-type distribution and 5.2251 (p-value 0.2650) for the S-type distribution, indicating an acceptable fit for both models, slightly better for the latter.

7 Conclusions

Two methods have been illustrated that allow to derive a bivariate discrete probability distribution from a continuous distribution, preserving the expression of the joint probability density function (first method) or the joint survival function (second method) of the original stochastic model. While the first method may not ensure a closed analytical form for the expression of the joint probability mass function of the discrete stochastic model, the second method always leads to an analytical expression, provided that the original joint survival function has one. Moreover, the second method is shown to enjoy some nice properties, for example maintaining also the marginal survival functions and the dependence structure (represented by the copula) between the two random components. This has a strong implication: both discrete marginal distributions depend only on the parameter of the corresponding continuous random component; this is not true for the first method, where the two marginal components generally turn out to depend on both marginal and dependence parameters of their parent bivariate continuous distribution. Apart from these general considerations, which let us to have a propensity for the second one, which of the two discretization technique performs better strongly depends on the real problem at study and on the criterion adopted for performance comparison.

Note that both discretization techniques have been designed in order to produce bivariate discrete distributions defined over $\mathbb{N} \times \mathbb{N}$, i.e., bivariate count variables. However, nothing prevents a more general design with a different countable support. For example, if we impose that X takes values in $\mathcal{X} = \{0, a, 2a, 3a, \dots\}$ and Y in $\mathcal{Y} = \{0, b, 2b, 3b, \dots\}$, with a and b positive real values, we have just to change Eqs. (3) and (4) as

$$p(x,y) = f(x,y)/\sum_{u \in \mathcal{X}} \sum_{v \in \mathcal{Y}} f(u,v), \quad x \in \mathcal{X}, y \in \mathcal{Y},$$

and

$$p(x,y) = S_{WZ}(x,y) + S_{WZ}(x+a,y+b) - S_{WZ}(x,y+b) - S_{WZ}(x+a,y), \quad x \in \mathcal{X}, y \in \mathcal{Y},$$

respectively.

Future research will investigate further methods of discretization in two dimensions that preserve other characteristics of the continuous stochastic model and their extension to higher dimension. We believe that given the increasing interest in the analysis of multivariate discrete data in many fields of the applied sciences, these techniques can be a useful tool for generating new stochastic models able to capture the features of real-world data. Moreover, they represent a theoretical and practical answer to the problem of approximating a (bivariate) continuous model by an appropriate discrete version, when the latter may strongly reduce the computational effort that the former would require.

Supplementary

`R` code used in the numerical example (Section 5) and in the applications to real data (Section 6) is avalaible at `https://tinyurl.com/ANOR-D-18-01277`.

References

Balakrishnan N, Lai C-D (2009). Distributions Expressed as Copulas. In: Continuous Bivariate Distributions. Springer, New York, pp 67-103

Barbiero A (2017). A bivariate geometric distribution with positive or negative correlation. In AIP Conference Proceedings 1906 (110003). AIP Publishing.

Barbiero A (2018a). A bivariate geometric distribution allowing for positive or negative correlation, Comm Statist Theory Methods, doi:10.1080/03610926.2018.1473428

Barbiero A (2018b). Properties and inferential issues of a bivariate version of the geometric distribution. In: 9th International Workshop on Applied Probability (IWAP 2018), Budapest, 18-21 June 2018.

Barbiero A (2019). Properties and estimation of a bivariate geometric model with locally constant failure rates, Ann Oper Res:1-20, doi 10.1007/s10479-019-03165-7

Barnett V (1980). Some bivariate uniform distributions. Comm Statist Theory Methods 9(4):453-461.

Bebbington M, Lai CD, Wellington M, Zitikis R (2012). The discrete additive Weibull distribution: A bathtub-shaped hazard for discontinuous failure data. Reliab Eng Syst Safe 106:37-44.

Bracquemond C, Gaudoin O (2003). A survey on discrete life time distributions. Int J Reliabil Qual Saf Eng 10:69-98

Chakraborty S (2015). Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. Journal of Statistical Distributions and Applications 2:6 `https://doi.org/10.1186/s40488-015-0028-6`

de Araujo Rodrigues J, Silva APCM, Chaves LM, Castellares F. (2011). Uma Nova Distribuicão Exponencial Bivariada Construída A Partir Da Copula De Gumbel-Barnett Com Uma Aplicação Em Modelagem De Precipitações Pluviométricas. Rev Bras Biom 29(4):583-98.

Farlie DJG (1960). The performance of some correlation coefficient for a general bivariate distribution. Biometrika 47:307-323.

Gumbel EJ (1960). Bivariate exponential distributions. J Amer Statist Assoc 55(292):698-707.

Inusah S, Kozubowski TJ (2006). A discrete analogue of the Laplace distribution. J Statist Plan Inference, 136(3):1090-1102.

Kemp AW (1997). Characterizations of a discrete normal distribution. J Stat Plan Inference 63(2):223-229.

Kotz S, Lumelskii Y, Pensky M (2003). The stress-strength model and its generalizations: theory and applications. World Scientific, New York.

Kozubowski TJ, Inusah, S (2006). A skew Laplace distribution on integers. Ann Inst Statist Math 58(3):555-571.

Johnson ME (1987). Multivariate Statistical Simulation, Wiley, New York

Lai CD (2006). Constructions of discrete bivariate distributions. In: Balakrishnan N, Castillo E, Sarabia Alegria JM (eds) Advances in Distribution Theory, Order Statistics, and Inference, pp 29-58.

Lai CD (2013). Issues concerning constructions of discrete lifetime models. Qual Technol Quant Manag 10(2):251-262

Loehlin JC (2004) Latent variable models: an introduction to factor, path, and structural equation analysis, 4th ed. Lawrence Erlbaum Associates, Mahwah, New Jersey

McGilchrist CA, Aisbett CW (1991). Regression with frailty in survival analysis. Biometrics 47:461-466

McNeil A, Frey R, Embrechts P (2005). Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton.

Mirhosseini SM, Amini M, Kundu D, Dolati A (2015). On a new absolutely continuous bivariate generalized exponential distribution. Stat Methods Appl 24:61-83

Mitchell C, Paulson A (1981). A new bivariate negative binomial distribution, Naval Res Logistics Quart 28:359-374.

Nakagawa T, Osaki S (1975). The discrete Weibull distribution. IEEE Trans. Reliability 24(5):300-301

Piperigou V (2009). Discrete distributions in the extended FGM family. J Stat Plan Inference 139(11): 3891-3899.

Roy D (1993). Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. J Multivariate Anal 46(2):362-373.

Roy D (2003). The discrete normal distribution. Comm Statist Theory Methods 32(10):1871-1883

Roy D, Dasgupta T (2001). A discretizing approach for evaluating reliability of complex systems under stress-strength model, IEEE Tran Reliab 50(2):145-150

Ruszczyński A, Shapiro A (2003). Stochastic Programming In: Ruszczyński A, Shapiro A (eds) Handbook in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, pp 1-64

Sklar A (1959). Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8:229-231.

Tancrez JS, Chevalier P, Sema P (2011). Probability masses fitting in the analysis of manufacturing flow lines, Ann Oper Res 182:163-191