AIP Conference Proceedings

Discrete analogs of a bivariate Pareto distribution

Cite as: AIP Conference Proceedings **2357**, 080009 (2022); https://doi.org/10.1063/5.0080639
Published Online: 09 May 2022

Alessandro Barbiero and Asmerilda Hitaj

Discrete Analogs of a Bivariate Pareto Distribution

Alessandro Barbiero[1,a)] and Asmerilda Hitaj[2,b)]

[1]*Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan (Italy)*
[2]*Department of Economics, Università degli Studi dell'Insubria, via Monte Generoso 71, 21100 Varese (Italy)*

[a)]Corresponding author: alessandro.barbiero@unimi.it
[b)]asmerilda.hitaj@uninsubria.it

Abstract. In many real-world problems, the phenomena of interest are continuous in nature and modeled through well-established continuous probability distributions, but it often occurs that observed values are actually discrete and then it would be more appropriate to use a (multivariate) discrete distribution generated from the underlying continuous model, which preserves one or more of its important features. In this work, we illustrate the genesis and properties of two bivariate discrete distributions that can be derived as discrete counterparts to a bivariate continuous Pareto distribution. The two discrete probability distributions preserve the expression of either 1) the joint density function or 2) the joint survival function of the parent distribution at each pair of non-negative integers. Their joint and marginal probability mass functions are derived and compared; the expressions for their bivariate failure rate vectors are also obtained. This study reveals how the second discrete analog is easier to handle with respect to the first one, whose expression of the probability mass function involves the Riemann zeta function.

INTRODUCTION

In many real-world applications, the phenomena of interest are continuous in nature and modeled through continuous probability distributions, but it often happens that their observed values are actually discrete and then it would be more appropriate to use a (multivariate) discrete distribution generated from the underlying continuous model, which preserves one or more of its important features.

If we consider a bivariate continuous random variable (rv) (W, Z), with joint probability density function (pdf) $f(w, z)$, cumulative distribution function (cdf) $F_{WZ}(w, z) = P(W \leq w, Z \leq z)$, and survival function (sf) $S_{WZ}(w, z) = P(W \geq w, Z \geq z)$, we may want to discretize it into a bivariate discrete rv (X, Y), with joint probability mass function (pmf) $p(x, y)$, cdf $F_{XY}(x, y)$, and joint 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., only in this region $f(w, z)$ is non-zero; we also assume that there $f(w, z)$ is finite.

Two methods have been proposed 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 pdf, or 2) the joint sf [1]. They are named D-type and S-type analogs, as they preserve the expression of the joint Density and Survival function, respectively. They have been then extended to the multivariate case by [2].

Method I: Bivariate Discretization Preserving the Joint PDF

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

$$p(x, y) = f_{WZ}(x, y) / \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} f_{WZ}(i, j), \quad x \in \mathbb{N}, y \in \mathbb{N}. \tag{1}$$

The distribution generated using this technique may not always have a compact closed form due to the double infinite summation acting as a normalizing constant at the denominator of Equation (1).

National Conference on Advances in Applied Sciences and Mathematics
AIP Conf. Proc. 2357, 080009-1–080009-5; https://doi.org/10.1063/5.0080639
Published by AIP Publishing. 978-0-7354-4189-7/$30.00

Method II: Bivariate Discretization Preserving the Joint SF

Given a bivariate continuous rv (W, Z) with sf $S_{WZ}(w,z)$, its S-type bivariate discrete analog is the rv (X, Y) with pmf

$$p(x,y) = S_{WZ}(x,y)+S_{WZ}(x+1,y+1)-S_{WZ}(x,y+1)-S_{WZ}(x+1,y) = \sum_{i=0}^{1}\sum_{j=0}^{1}(-1)^{i+j}S_{WZ}(x+i,y+j), \quad x,y \in \mathbb{N}. \quad (2)$$

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 analog also preserves the expression of the sfs of the marginal components at any non-negative integer.

In the next section, we will apply these two methods for constructing a discrete counterpart to a bivariate continuous Pareto distribution and compare the two resulting discrete distributions. Some final remarks and further research perspectives are addressed in the last section.

CONTRUCTION OF TWO DISCRETE ANALOGS OF A BIVARIATE PARETO DISTRIBUTION

Let us consider the bivariate continuous distribution with joint sf

$$S_{WZ}(w,z) = (1+w+z)^{-c}, \quad w,z \in \mathbb{R}^+,$$

with $c > 0$, whose joint pdf is given by

$$f_{WZ}(w,z) = c(c+1)(1+w+z)^{-c-2}, \quad (3)$$

which is known as the bivariate Pareto distribution [3]. The marginal pdf of W is given by $f_W(w) = c(1+w)^{-(c+1)}$; the marginal cdf by $F_W(w) = 1-(w+1)^{-c}$; the marginal sf by $S_W(w) = (w+1)^{-c}$: hence, W is a Pareto type II (Lomax) distribution [4]; same arguments hold for Z. It can be shown that the unique copula C capturing the dependence structure between W and Z is the survival copula associated to the Clayton copula with parameter c [3, p.90]. Henceforth, we will consider the simple case $c = 1$; for this parameter choice, marginal expected values and variances are undefined as well as Pearson's linear correlation.

The first discretization technique, through Equation (1), yields the discrete rv (X, Y) with joint pmf

$$p(x,y) = \frac{2C^{-1}}{(1+x+y)^3}, \quad x,y \in \mathbb{N}, \quad (4)$$

where

$$C = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}2(1+i+j)^{-3} = 2\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{1}{(i+j)^3} = 2\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{(i+j)^3}+\sum_{i=1}^{\infty}\frac{1}{i^3}\right) = 2(\zeta(2)-\zeta(3)+\zeta(3)) = 2\zeta(2) \approx 3.289868,$$

being $\zeta(s) = \sum_{n=1}^{\infty}1/n^s$, $s \in \mathbb{C}$, $\mathrm{Re}(s) > 1$, the Riemann zeta function. The marginal pmf of X can be computed as

$$p_X(x) = \sum_{j=0}^{\infty}p(x,j) = -C^{-1}\psi^{(2)}(x+1),$$

where $\psi^{(m)}(\cdot)$ is the polygamma function of order m, $\psi^{(m)}(x) = \frac{\mathrm{d}^{m+1}}{\mathrm{d}x^{m+1}}\log(\Gamma(x))$, with $\Gamma(x) = \int_0^{\infty}t^{x-1}e^{-t}\mathrm{d}t$. The joint survival function can be written as

$$S_{xy}(x,y) = \sum_{i=x}^{\infty}\sum_{j=y}^{\infty}p(i,j) = \sum_{i=x}^{\infty}\sum_{j=y}^{\infty}\frac{2C^{-1}}{(1+i+j)^3} = 2C^{-1}\sum_{k=1}^{\infty}\frac{k}{(k+x+y)^3}.$$

The second discretization technique, through Equation (2), provides the following expression for the joint pmf of the rv (X, Y):

$$p(x,y) = \frac{1}{1+x+y}+\frac{1}{3+x+y}-\frac{2}{2+x+y} = \frac{2}{(1+x+y)(2+x+y)(3+x+y)}. \quad (5)$$

It is easy to derive the marginal pmf of X, by directly exploiting the telescoping nature of the joint pmf (5):

$$p_X(x) = \sum_{j=0}^{\infty} p(x, j) = \frac{1}{1+x} - \frac{1}{2+x} = \frac{1}{(1+x)(2+x)}.$$

The pmf of X can be actually derived as the discrete analog (through preservation of the sf) of a Pareto II (Lomax) distribution, since $S_W(x) = 1/(1+x)$ and $S_W(x) - S_W(x+1) = 1/(1+x) - 1/(2+x)$; it is a special case of the discrete Burr distribution introduced in [5], with parameters $\alpha = \beta = 1$. So a bivariate continuous distribution with Pareto margins is here discretized into a bivariate discrete distribution with (discrete) Pareto margins.

In Table 1, the joint pmfs of the two discrete analogs (4) and (5) of the bivariate Pareto distribution are (partially) reported (only the values $x, y = 0, 1, 2, 3$ are considered for the sake of brevity).

TABLE 1. Joint pmfs of the D-type (left) and S-type (right) discrete analogs of the bivariate Pareto distribution (3), with $c = 1$

x, y	0	1	2	3		$P(X = x)$
0	.6079	.0760	.0225	.0095	···	.7307
1	.0760	.0225	.0095	.0049	···	.1228
2	.0225	.0095	.0049	.0028	···	.0468
3	.0095	.0049	.0028	.0018	···	.0243
···	···	···	···	···	···	···
$P(Y = y)$	.7307	.1228	.0468	.0243	···	1

x, y	0	1	2	3		$P(X = x)$
0	.3333	.0833	.0333	.0167	···	.5000
1	.0833	.0333	.0167	.0095	···	.1667
2	.0333	.0167	.0095	.0060	···	.0833
3	.0167	.0095	.0060	.0040	···	.0500
···	···	···	···	···	···	···
$P(Y = y)$	.5000	.1667	.0833	.0500	···	1

We will derive now the expression of a reliability-related function for the bivariate Pareto distribution. For a bivariate continuous rv (W, Z) we can define a bivariate failure rate vector [see e.g. 6, 7] as

$$(r_1, r_2) := \left(-\frac{\partial}{\partial w}\log S_{WZ}(w, z), -\frac{\partial}{\partial z}\log S_{WZ}(w, z)\right).$$

The bivariate failure rate vector for the bivariate Pareto distribution is thus given by

$$(r_1, r_2) = \left(\frac{1}{1+w+z}, \frac{1}{1+w+z}\right).$$

For a bivariate discrete rv (X, Y) we can define a bivariate failure rate vector [8] as

$$(\lambda_1, \lambda_2) = \left(\frac{P(X = x, Y \geq y)}{S_{XY}(x, y)}, \frac{P(X \geq x, Y = y)}{S_{XY}(x, y)}\right).$$

λ_1 (λ_2) can be regarded as a discrete version of r_1 (r_2). In fact, r_1 can be rewritten as $-\frac{\partial S_{WZ}(w,z)}{\partial w}\Big/S_{WZ}(w, z)$, whose discrete analog would be $-\frac{S_{XY}(x+1,y)-S_{XY}(x,y)}{S_{XY}(x,y)} = \frac{P(X=x,Y\geq y)}{P(X\geq x,Y\geq y)}$, i.e., λ_1. Thus, for the D-type discrete analog (4) we have:

$$\begin{aligned}\lambda_1(x, y) = \lambda_2(x, y) &= \sum_{t=y}^{\infty}\frac{1}{(1+x+t)^3}\Big/\sum_{k=1}^{\infty}\frac{k}{(k+x+y)^3} = \sum_{k=1}^{\infty}\frac{1}{(k+x+y)^3}\Big/\sum_{k=1}^{\infty}\frac{k}{(k+x+y)^3}\\ &= -\psi^{(2)}(1+x+y)/[2\psi^{(1)}(1+x+y) + (x+y)\psi^{(2)}(1+x+y)].\end{aligned}$$

For the S-type discrete analog (5) we have that

$$P(X = x, Y \geq y) = \sum_{t=y}^{\infty} p(x, t) = \sum_{t=y}^{\infty}\frac{1}{1+x+t} + \frac{1}{3+x+t} - \frac{1}{2+x+t} = \frac{1}{1+x+y} - \frac{1}{2+x+y} = \frac{1}{(1+x+y)(2+x+y)}$$

and then

$$\lambda_1(x, y) = \lambda_2(x, y) = \frac{1}{2+x+y}.$$

The graphs of the identical components of the bivariate failure rate vectors for the bivariate Pareto distribution and its two discrete analogs are displayed in Figure 1 as a function of $s = x + y$. Since the following asymptotic formulas hold for $s \to \infty$ [see 9, Equations 6.4.12-6.4.13]:

$$\psi^{(1)}(s) \sim \frac{1}{s} + \frac{1}{2s^2} + \frac{1}{6s^3} + \dots, \qquad \psi^{(2)}(s) \sim -\frac{1}{s^2} - \frac{1}{s^3} - \frac{1}{2s^4} + \dots,$$

it can be easily noticed that the bivariate failure rate for the D-type discrete analog asymptotically tends to the bivariate failure rate of its continuous counterpart: $\lim_{s\to\infty} \lambda_1(x, y)/r_1(x, y) = 1$. Similarly, for the S-type discrete analog we have $\lim_{s\to\infty} \lambda_1(x, y)/r_1(x, y) = 1$. This result can be also observed in Figure 1.

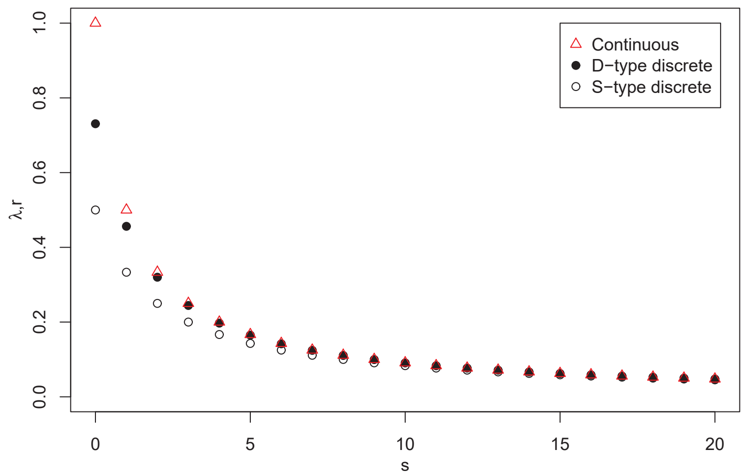

FIGURE 1. Bivariate failure rates for the bivariate Pareto distribution ($r_1(x, y) = r_2(x, y)$) and its two discrete analogs ($\lambda_1(x, y) = \lambda_2(x, y)$) as a function of $s = x + y$

FURTHER RESEARCH

In this work, we constructed two discrete analogs of a bivariate continuous Pareto distribution when its parameter c equals 1, and compared them in terms of the joint probability mass function and bivariate failure rates. Further research will address discretization of the same distribution for different values of c. However, we remark that while the discretization technique based on the joint sf is very straightforward, at least when the sf of the original distribution is available in a closed-form expression, the technique based on the joint pdf is often unable to lead to a closed-form expression for the bivariate pmf of the discrete counterpart, as occurs for the distribution studied here.

ACKNOWLEDGMENTS

The first author acknowledges the financial support to the present research by the University of Milan (Piano di Sostegno alla Ricerca 2018-Linea 2A).

REFERENCES

[1] A. Barbiero, "Discrete analogues of continuous bivariate probability distributions". *Ann Oper Res*, https://doi.org/10.1007/s10479-019-03388-8 (2020).

[2] M. Wiegand, S. Nadarajah, Y. Zhang, "Discrete analogues of continuous multivariate probability distributions". *Ann Oper Res*, https://doi.org/10.1007/s10479-020-03633-5 (2020)

[3] N. Balakrishnan and C.-D. Lai, "Distributions Expressed as Copulas," in *Continuous Bivariate Distributions* (Springer, New York, 2009), pp 67–103.

[4] K.S. Lomax, "Business Failures; Another example of the analysis of failure data", *J Am Stat Assoc* **49**, 847–852 (1954).
[5] H. Krishna and P.S. Pundir, "Discrete Burr and discrete Pareto distributions", *Stat Methodology* **6**, 177–188 (2009).
[6] M.E. Johnson, *Multivariate Statistical Simulation* (Wiley, New York, 1987).
[7] J. Navarro, "Characterizations using the bivariate failure rate function", *Stat Prob Letters* **78(12)**, 1349–1354 (2008).
[8] D. Roy, "Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution". *J Multivariate Anal* **46(2)**, 362–373 (1993).
[9] M. Abramowitz and M.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).