Discrete analogues of continuous multivariate probability distributions

Martin Wiegand and Saralees Nadarajah

today

Abstract

In this note we introduce a multivariate extension to the bivariate discrete analogoue to continous probability distributions proposed in [1]. We furthermore show that identical properties of the bivariate case also hold for n-variate discrete analogues.

1 Introduction

2 Discrete Analogues

In this section we will derive the discrete analogue of an n-variate continuous random variable. We will do so in accordance with the two approaches originally proposed by Barbiero in [1]. We begin with the first method which is based on the continuous denisty function (pdf) of the continuous parent random variable.

2.1 Method 1

Theorem 1. Let $(W_1, \ldots, W_n)$ be a continuous n-variate random variable with corresponding pdf $f_{(W_1,\ldots,W_n)}$, then its D-type discrete analogue random variable is $(X_1, \ldots, X_n)$ with the probability mass function as follows:

$$p_{(X_1,\ldots,X_n)} = \frac{f_{(W_1,\ldots,W_n)}(x_1,\ldots,x_n)}{\sum_{w_1\in\mathbb{N}}\cdots\sum_{w_n\in\mathbb{N}} f_{(W_1,\ldots,W_n)}(w_1,\ldots,w_n)} \qquad (1)$$

Proof. Deriving the two main properties of the pmf is trivial. As the pdf is always non-negative $f_{(W_1,\ldots,W_n)}(x_1,\ldots,x_n) \geq 0$ for all $(x_1,\ldots,x_n) \in \mathbb{N}^n$ both denominator and numeratorof the pmf are non-negative and the property follows. The second property may be derived as follows:

$$\sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{(X_1,\ldots,W_n)}(i_1,\ldots,i_n) = \frac{\sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} f_{(W_1,\ldots,W_n)}(x_1,\ldots,x_n)}{\sum_{w_1 \in \mathbb{N}} \cdots \sum_{w_n \in \mathbb{N}} f_{(W_1,\ldots,W_n)}(w_1,\ldots,w_n)}$$

$$= 1 \tag{2}$$

$\square$

2.2 Method 2

Theorem 2. Let $(W_1,\ldots,W_n)$ be a multivariate continuous random variable with the survival function $S_{W_1,\ldots,W_n}(w_1,\ldots,w_n)$ and CDF $F_{W_1,\ldots,W_n}(w_1,\ldots,w_n)$ then the S-type multivariate discrete andalogue random variable $(X_1,\ldots,X_n)$ possess the pmf below:

$$p(x_1,\ldots,x_n) = \sum_{k=0}^{n}(-1)^{n-k}\sum_{t=0}^{\binom{n}{k}} F_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}}) \tag{3}$$

where $\mathbf{x} = (\mathbf{x_1},\ldots,\mathbf{x_n})$ and ρ_t^k with $t = 1,\ldots,\binom{n}{k}$ are the permutations of a vector of length n with k values of $x_i = 1$ and the remaing $n - k$ values zero. Alternatively we can define the analogous formulation using the survival function below:

$$p(x_1,\ldots,x_n) = \sum_{k=0}^{n}(-1)^{n-k+1}\sum_{t=0}^{\binom{n}{k}} S_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}}) \tag{4}$$

Proof. We have to show the two basic properties of the pmf, e.g. that $p(x_1,\ldots,x_n) \geq 0$, and that $\sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} = 1$. We begin with property one analogously to the proof provided for the bivariate case by Barbieri in [1]:

$$p(x_1,\ldots,x_n) = \sum_{k=0}^{n}(-1)^{n-k}\sum_{t=0}^{\binom{n}{k}} F_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}}) \tag{5}$$

$$= \mathbb{P}\left[x_1 < W_1 < x_1 + 1, \ldots, x_n < W_n < x_n + 1\right] \tag{6}$$

$$\geq 0 \tag{7}$$

Similarly we can derive the second property:

$$\sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} p(x_1, \ldots, x_n)$$

$$= \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \sum_{k=0}^{n} (-1)^{n-k} \sum_{t=0}^{\binom{n}{k}} F_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}})$$

$$= \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \sum_{k=0}^{n} (-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} S_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}})$$

$$= \sum_{x_2=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \sum_{x_1=0}^{\infty} (-1)^{n+1} S(x_1 + 1, x_2 + 1, \ldots, x_n + 1) \tag{8}$$

$$+ (-1)^n S_{(W_1,\ldots,W_n)}(x_1, x_2 + 1, \ldots, x_n + 1)$$
$$+ (-1)^n S_{(W_1,\ldots,W_n)}(x_1 + 1, x_2, x_3 + 1, \ldots, x_n + 1) \tag{9}$$
$$+ (-1)^{n-1} S_{(W_1,\ldots,W_n)}(x_1, x_2, x_3 + 1, \ldots, x_n + 1)$$
$$+ \cdots + S_{(W_1,\ldots,W_n)}(x_1, \ldots, x_n)$$

$$= \sum_{x_2=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} S_{(W_1,\ldots,W_n)}(0, x_2 + 1, \ldots, x_n + 1)$$
$$- S_{(W_1,\ldots,W_n)}(0, x_2, x_3 + 1, \ldots, x_n + 1)$$
$$+ \cdots + S_{(W_1,\ldots,W_n)}(0, \ldots, x_n)$$

$$\vdots$$

$$= S_{(W_1,\ldots,W_n)}(0, \ldots, 0) = 1 \tag{10}$$

For each x_i the telescope sum properties of the pmf lets us reduce the 2^n sum terms by half, while leaving the respective variable with a zero value. Repeating this n times leaves us with the n-dimensional survival function at $\mathbf{x} = (\mathbf{0}, \ldots, \mathbf{0})$. $\square$

Similarly we may derive the same corollaries as with the bivariate case from the original manuscript [1].

Proposition 3. Let $(X_1, \ldots, X_n)$ be the S-type multivariate discrete random variable and $(W_1, \ldots, W_n)$ its continuous parent random variable. Then two properties hold, firstly for the survival functions of both random variables:

$$S_{(X_1,\ldots,X_n)}(h_1, \ldots, h_n) = S_{(W_1,\ldots,W_n)}(h_1, \ldots, h_n) \tag{11}$$

where $h_i \in \mathbb{N}^n$ for $i = 1, \ldots, n$. Secondly the marginal survival functions are preserved.

$$S_{X_i}(h) = S_{W_i}(h) \qquad \forall i \in 1, \ldots, n \tag{12}$$

3

Proof. The first property follows by the same logic as used for Theorem 2:

$$\sum_{x_1=h_1}^{\infty} \cdots \sum_{x_n=h_2}^{\infty} p(x_1,\ldots,x_n)$$

$$= \sum_{x_1=h_1}^{\infty} \cdots \sum_{x_n=h_n}^{\infty} \sum_{k=0}^{n} (-1)^{n-k} \sum_{t=0}^{\binom{n}{k}} F_{(W_1,\ldots,W_n)}(\mathbf{x}+\rho_{\mathbf{t}}^{\mathbf{k}})$$

$$= \sum_{x_1=h_1}^{\infty} \cdots \sum_{x_n=h_n}^{\infty} \sum_{k=0}^{n} (-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} S_{(W_1,\ldots,W_n)}(\mathbf{x}+\rho_{\mathbf{t}}^{\mathbf{k}})$$

$$= \sum_{x_2=h_2}^{\infty} \cdots \sum_{x_n=h_n}^{\infty} \sum_{x_1=h_1}^{\infty} (-1)^{n+1} S(x_1+1, x_2+1,\ldots,x_n+1) \qquad (13)$$

$$+ (-1)^{n} S_{(W_1,\ldots,W_n)}(x_1, x_2+1,\ldots,x_n+1)$$

$$+ (-1)^{n} S_{(W_1,\ldots,W_n)}(x_1+1, x_2, x_3+1,\ldots,x_n+1) \qquad (14)$$

$$+ (-1)^{n-1} S_{(W_1,\ldots,W_n)}(x_1, x_2, x_3+1,\ldots,x_n+1)$$

$$+ \cdots + S_{(W_1,\ldots,W_n)}(x_1,\ldots,x_n)$$

$$= \sum_{x_2=h_2}^{\infty} \cdots \sum_{x_n=h_n}^{\infty} S_{(W_1,\ldots,W_n)}(h_1, x_2+1,\ldots,x_n+1)$$

$$- S_{(W_1,\ldots,W_n)}(h_1, x_2, x_3+1,\ldots,x_n+1)$$

$$+ \cdots + S_{(W_1,\ldots,W_n)}(h_1,\ldots,x_n)$$

$$\vdots$$

$$= S_{(W_1,\ldots,W_n)}(h_1,\ldots,h_n) \qquad (15)$$

From the first property we may simply derive the marginal properties as follows:

$$S_{X_i}(h) = S_{(X_1,\ldots,X_n)}(0,\ldots,h,\ldots,0) = S_{(W_1,\ldots,W_n)}(0,\ldots,h,\ldots,0) = S_{W_i}(h) \qquad (16)$$

$\square$

Proposition 4. Let W_i for $i = 1,\ldots,n$ be the marginal components of the continuous n-variate random variable $(W_1,\ldots,W_n)$ then X_i the marginal components of the discrete n-variate S-type random variable $(X_1,\ldots,X_n)$ are the discrete analogues for all $i = 1,\ldots,n$.

Proof. For all $i = 1,\ldots,n$ the proof follows analogously:

$$p_{X_i}(h) = S_{(X_i}(h) - S_{X_i}(h+1) = S_{W_i}(h) - S_{W_i}(h+1). \qquad (17)$$

4

Based on Barbiero (univariate result 2.2 [1]) this indicates that the marginal discrete distribution of X_i is the discrete analogue of the continuous parent distribution of the marginal random variable W_i, for all $i = 1, \ldots, n$. Similarly we derive for the marginal CDF:

$$F_{X_i} = 1 - \mathbb{P}[X \geq x + 1] = 1 - S_{X_i}(x + 1) \stackrel{\text{Prop. 3}}{=} 1 - S_{W_i}(x + 1) = F_{W_i}(x + 1) \tag{18}$$

for all $i = 1, \ldots, n$.

$\square$

Based on the previous result we may derive the next property of the n-variate discrete analogue.

Proposition 5. Let the n-variate conitnuous random variable $(W_1, \ldots, W_n)$ have the copula $\mathcal{C}$, then the S-type n-variate discrete analogue $(X_1 \ldots, X_n)$ possess the same copula $\mathcal{C}$ on the same range $Range(F_{X_1}) \times \cdots \times Range(F_{X_n})$.

Proof. By [2] it holds that:

$$F_{(W_1, \ldots, W_n)}(w_1, \ldots, w_n) = \mathcal{C}(F_{W_1}(w_1), \ldots, F_{W_n}(w_n)) \tag{19}$$

We may then follow for $(h_1 \ldots, h_n) \in \mathbb{N}^n$:

$$F_{(X_1\ldots,X_n)}(h_1,\ldots,h_n) = \sum_{i=1}^{n}\mathbb{P}[X_i \leq h_i] - \sum_{i<j}^{n} P[X_i \leq h_i, X_j < h_j]$$

$$+ \sum_{i<j<k}^{n}\mathbb{P}[X_i \leq h_i, X_j \leq h_j, X_k \leq h_k]$$

$$+ \cdots + (-1)^n\mathbb{P}[X_1 \leq h_1,\ldots,X_n \leq h_n]$$

$$= \sum_{i=1}^{n} S_{X_i}(h_i+1) - \sum_{i<j}^{n} S_{X_i,X_j}(h_i+1,h_j+1)$$

$$+ \sum_{i<j<k}^{n} S_{(X_i,X_j,X_k)}()$$

$$+ \cdots + (-1)^n S_{(X_1,\ldots,X_n)}(h_1+1,\ldots,h_n+1)$$

$$= \sum_{i=1}^{n} S_{W_i}(h_i+1) - \sum_{i<j}^{n} S_{W_i,W_j}(h_i+1,h_j+1)$$

$$+ \sum_{i<j<k}^{n} S_{(W_i,W_j,W_k)}()$$

$$+ \cdots + (-1)^n S_{(W_1,\ldots,W_n)}(h_1+1,\ldots,h_n+1)$$

$$= F_{(W_1,\ldots,W_n)}(h_1+1,\ldots,h_n+1)$$

$$\overset{Eq.19}{=} \mathcal{C}(F_{W_1}(h_1+1),\ldots,F_{W_n}(h_n+1))$$

$$\overset{Prop.4}{=} \mathcal{C}(F_{X_1}(h_1),\ldots,F_{X_n}(h_n)) \tag{20}$$

$\square$

With the previous Propositions 3 - 5 in place, we may derive the final property. The main results of this proposition is the identity of the discrete analogue rv $(X_1,\ldots,X_n)$ of the continuous n-variate random variable $(W_1,\ldots,W_n)$ and it's rounded counterpart $(\lfloor W_1 \rfloor,\ldots,\lfloor W_n \rfloor)$

Proposition 6. Let $(W_1,\ldots,W_n)$ be an n-variate continuous random variable. Then its S-type n-variate discrete analogue random variable $(X_1,\ldots,X_n)$ is of the same distribution as $(\lfloor W_1 \rfloor,\ldots,\lfloor W_n \rfloor)$

Proof. We know that $(\lfloor W_1 \rfloor,\ldots,\lfloor W_n \rfloor) \in \mathbb{N}^n$. We may use Proposition 3 to derive the following for $(h_1,\ldots,h_n) \in \mathbb{N}^n$.

$$\mathbb{P}[\lfloor W_1 \rfloor = h_1,\ldots,\lfloor W_n \rfloor = h_n] = \mathbb{P}[h_1 \leq W_1 < h_1+1,\ldots,h_n \leq W_n < h_n+1]$$

$$\overset{Prop.3}{=} \sum_{k=0}^{n}(-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} S_{(W_1,\ldots,W_n)}(\mathbf{x}+\rho_{\mathbf{t}}^{\mathbf{k}})$$

$$= p_{(X_1,\ldots,X_n)}(h_1,\ldots,h_n) \tag{21}$$

6

$\square$

3   Applications

To illustrate the applications of the results in the previous section we provide some examples.

Example 3.1. Let $(W_1, \ldots, W_n)$ be of n-variate normal distribution, with parameters $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ and a positive semi-definite matrix $\Sigma \in \mathbb{R}^{n \times n}$ on its support $\mathbb{R}^n$. The pmf of the discrete analogue random variable $(X_1, \ldots, X_n)$ can then be written by the first method presented as:

$$
\begin{aligned}
p_{(X_1,\ldots,X_n)} &= \frac{(2\pi)^{-n/2} det(\Sigma)^{-1/2} e^{-1/2(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} (2\pi)^{-n/2} det(\Sigma)^{-1/2} e^{-1/2(\mathbf{i}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{i}-\mu)}} \\
&= \frac{e^{-1/2(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} e^{-1/2(\mathbf{i}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{i}-\mu)}} \\
&= \frac{1}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} e^{1/2((\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)-(\mathbf{i}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-1}(\mathbf{i}-\mu))}}
\end{aligned}
\tag{22}
$$

where $\mathbf{i} = (\mathbf{i_1}, \ldots, \mathbf{i_n}) \in \mathbb{N}_\mathbf{0}^\mathbf{n}$

A second example can easily be derived by the second discretisation approach.

Example 3.2. Consider $(W_1, \ldots, W_n)$ to be of n-variate Pareto distribution type II, with parameters $\sigma_i, \alpha > 0$ and $\mu_i = 0$ for all $i = 1, \ldots, n$ for $w_i > 0$ for all $i = 1, \ldots, n$. The pmf of the discrete analogue random variable $(X_1, \ldots, X_n)$ is in accordance with the second method constructed as follows:

$$
\begin{aligned}
p_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) &= \sum_{k=0}^{n} (-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} S_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_\mathbf{t}^\mathbf{k}) \\
&= \sum_{k=0}^{n} (-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} \left(1 + \frac{\sum_{i=1}^{n}(x_i + \rho_{i,t}^k)}{\sum_{i=1}^{n} \sigma_i}\right)^{-\alpha} \\
&= \sum_{k=0}^{n} (-1)^{n-k+1} \sum_{t=0}^{\binom{n}{k}} \left(1 + \frac{\sum_{i=1}^{n} x_i + k}{\sum_{i=1}^{n} \sigma_i}\right)^{-\alpha} \\
&= 2^n \sum_{k=0}^{n} (-1)^{n-k+1} \left(1 + \frac{\sum_{i=1}^{n} x_i + k}{\sum_{i=1}^{n} \sigma_i}\right)^{-\alpha}
\end{aligned}
\tag{23}
$$

Lastly we close this with an example of the second method in its variation utilising the CDF rather than the survival function of the continuous parent random variable.

Example 3.3. Let $(W_1, \ldots, W_n)$ be of the modified Clayton's copula distribution with parameter $\alpha > 0$ on the support $(w_1, dots, w_n) \in \mathbb{R}^{n,+}$. The corresponding CDF F is the following:

$$F(w_1, \ldots, w_n) = \left[\sum_{i=1}^{n} \exp(-\alpha w_i) - n + 1 \right]^{-\frac{1}{\alpha}}$$

We may then follow the pmf of the discrete analogoue random variable $(X_1, \ldots, X_n)$ by the CDF formulation of the second discretisation approach.

$$
\begin{aligned}
p_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) &= \sum_{k=0}^{n}(-1)^{n-k} \sum_{t=0}^{\binom{n}{k}} F_{(W_1,\ldots,W_n)}(\mathbf{x} + \rho_{\mathbf{t}}^{\mathbf{k}}) \\
&= \sum_{k=0}^{n}(-1)^{n-k} \sum_{t=0}^{\binom{n}{k}} \left[\sum_{i=1}^{n} \exp(-\alpha x_i + \rho_{i,t}^{k}) - n + 1 \right]^{-\frac{1}{\alpha}}
\end{aligned}
\tag{24}
$$

4   Conclusion

In this note we have extended the original bivariate results of Barbiero to their multivariate formulation. Starting from the original proofs we were able the derive the same properties of the discrete analogue for arbitrary dimensions. We close this paper by providing application examples.

References

[1] A. Barbierio,
Discrete analogues of continuous bivariate probability distributions
Annals of Operations Research. (2019)

[2] A. Sklar,
Fonctions de repartition a n dimensions et leurs marges.
Publications de l'institute de Statistique de l'Universite de Paris, 8, p.229-231. (1959)