Research Article   **Open Access**

Gery Geenens*

Copula modeling for discrete random vectors

Abstract: Copulas have now become ubiquitous statistical tools for describing, analysing and modelling dependence between random variables. Sklar's theorem, "the fundamental theorem of copulas", makes a clear distinction between the continuous case and the discrete case, though. In particular, the copula of a discrete random vector is not fully identifiable, which causes serious inconsistencies. In spite of this, downplaying statements may be found in the related literature, where copula methods are used for modelling dependence between discrete variables. This paper calls to reconsidering the soundness of copula modelling for discrete data. It suggests a more fundamental construction which allows copula ideas to smoothly carry over to the discrete case. Actually it is an attempt at rejuvenating some century-old ideas of Udny Yule, who mentioned a similar construction a long time before copulas got in fashion.

Keywords: Copula, discrete random vector, bivariate Bernoulli distribution, Yule's colligation coefficient, iterative proportional fitting

MSC: 62H05, 62H17, 62H20

1 Introduction

In Yule [53], one can read: "*Two association tables that are not directly comparable owing to the different proportions of A's and B's in the data from which the tables were compiled may be rendered directly comparable by multiplying the frequencies in rows and columns by appropriate factors,* [...] *reducing the original tables to some arbitrarily selected standard form*" (p. 588). The standard form that he recommends is the table whose margins have been made uniform. Likewise, in their extensive study of association coefficients in (2×2)-contingency tables, Goodman and Kruskal [21, p. 747] mentioned transforming all marginals to $1/2$ for facilitating interpretation. Later, Mosteller [36] developed: "*We might instead think of a contingency table as having a basic* nucleus *which describes its association and think of all tables formed by multiplying elements in rows and columns by positive numbers as forming an equivalence class – a class of tables with the same degree of association*" (p. 4). And: "*we might especially arrange the table to have uniform margins on each side in the case of a two-way table so as to get a clearer look at the association that is actually occurring*" (p. 6).

If one identifies two-way contingency tables with bivariate discrete distributions, then the above ideas have much in common with copulas: one tries to capture the dependence structure between the two variables apart from the marginal distributions by making these into uniforms, hence uninformative. The observation is notable, as it has been known at least since Marshall [32] that the notion of copula fits poorly in the discrete framework. Here 'copula' refers to the classical definition [7, Definition 1.3.1] which, in the bivariate case, reads:

Definition 1.1. *A bivariate copula C is a function from $\mathcal{I} \doteq [0,1]^2$ to $[0,1]$ defined as $C(u,v) = \mathbb{P}(U \le u, V \le v)$, where $U, V \sim \mathcal{U}_{[0,1]}$, the continuous uniform distribution on the unit interval.*

*Corresponding Author: Gery Geenens:** School of Mathematics and Statistics, UNSW Sydney, Australia, tel +61 2 938 57032, fax +61 2 9385 7123, E-mail: ggeenens@unsw.edu.au

Such copulas naturally arise in statistical modelling through the celebrated Sklar's theorem [50]:

Theorem 1.1 (Sklar). *Let F_{XY} be the distribution function of a bivariate random vector (X, Y), with marginal distribution functions F_X and F_Y. Then there exists a copula C such that, for all $(x, y) \in \mathbb{R}^2$,*

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)). \tag{1.1}$$

If F_X and F_Y are continuous, then C is unique; otherwise C is uniquely determined on $\mathrm{Ran}\, F_X \times \mathrm{Ran}\, F_Y$ only. Conversely, for any univariate distribution functions F_X and F_Y and any copula C, the function F_{XY} defined by (1.1) is a valid bivariate distribution function with marginals F_X and F_Y.

The popularity of copulas for dependence modelling largely follows from quotes like '*Copulas allow us to separate the effect of dependence from effects of the marginal distributions*'. Clearly, if C is unique, then it unequivocally characterises how the two marginals F_X and F_Y interlock for producing the joint behaviour of (X, Y), while staying ignorant of what those marginals are. For instance, if X and Y are independent ($X \perp\!\!\!\perp Y$) *and* if C is unique, then from (1.1) C must be the 'independence copula' (or 'product copula')

$$\Pi(u, v) = uv \qquad \forall (u, v) \in \mathcal{I}, \tag{1.2}$$

and this regardless of F_X and F_Y. It is often overlooked that the situation is this appealing only in the case of continuous margins. When there is no one-to-one correspondence between the joint distribution F_{XY} and the copula C, i.e., for X and/or Y discrete, the above argument falls apart.

Instrumental to copula ideas is the vector $(F_X(X), F_Y(Y))$. If X and Y are both continuous, then, through '*Probability Integral Transform*' (hereafter: PIT), $F_X(X)$ and $F_Y(Y)$ have uniform distributions $\mathcal{U}_{[0,1]}$, and the copula C is their joint distribution. Clearly one can plug any increasing transformations of X and/or Y into PIT with the same output. Hence copulas are invariant under increasing transformations of the margins [37, Theorem 2.4.3], that is, 'margin-free'. Any copula-based dependence measure, such as Kendall's or Spearman's correlations [37, Chapter 5], is then margin-free as well.

Now, in the case X and/or Y discrete, $\mathrm{Ran}\, F_X$ and/or $\mathrm{Ran}\, F_Y$ are just countable subsets of $[0, 1]$. The distributions of $F_X(X)$ and/or $F_Y(Y)$ are thus not $\mathcal{U}_{[0,1]}$, and their joint distribution cannot be a copula as described by Definition 1.1. It is actually a *subcopula*, i.e., a function satisfying the main structural properties of copulas but whose support is only a strict subset of $\mathcal{I}$ containing 0 and 1 [37, Definition 2.2.1]. Any such subcopula can be extended into a copula [37, Lemma 2.3.5]: the gaps in $\mathcal{I} \setminus (\mathrm{Ran}\, F_X \times \mathrm{Ran}\, F_Y)$ can be filled in a way preserving the properties of copulas; however there are uncountably many ways of doing so and C in (1.1) is not identifiable.

Such unidentifiability *does* cause serious inconsistencies, which [18] systematically investigated following preliminary warnings in [32], and questions the soundness of copula modelling for discrete data. This is discussed in Section 2, while a further invitation to calling current practice into question is offered in Section 3. In particular, we claim that the concept of 'copula' should be given a more fundamental meaning, not limited to Definition 1.1 but agreeing with it in the continuous case. The key is to apprehend copulas from a different perspective, a detailed discussion of which is given in Section 4.

Then we propose a construction which, rejuvenating Yule's, Goodman and Kruskal's and Mosteller's conceptions, allows a seamless extension of the main ideas of copula theory to the discrete case; first in the case of a bivariate Bernoulli distribution (Section 5) and then gradually generalising it to bivariate discrete distributions with finite support (Section 6) and finally with infinite support (Section 8). Along the way, we bridge the gap between the copula literature and methods of contingency tables analysis. In particular, the role of the odds-ratio is reinforced and Yule's colligation coefficient establishes itself as the appropriate dependence parameter in a discrete copula-like setting; algebraic and geometric representations of contingency tables are leveraged; parallels between discrete copula modelling and the problem of matrix scaling are drawn; and a novel visual display of bivariate discrete distributions ('*confetti plot*') is proposed. New parametric models of dependence between discrete random variables are also introduced in Section 7.

2 Copulas on discrete distributions

Most reasons which make copula modelling attractive and effective in the continuous case, break up in the discrete case: "*everything that can go wrong, will go wrong*" [10, p. 641]. A major downside is that copulas lose their margin-free nature when applied to discrete random vectors – whereas the whole copula methodology came into being in the first place for exploiting the benefits of margin-freeness [47]. In particular, in the discrete case, copula-based measures of dependence (e.g. Kendall's or Spearman's) are margin-dependent [Proposition 2.3 in 18, 32, Section 4.2]. Worse, a given copula model *per se* may or may not be intrinsically meaningful or even compatible with some marginals [12, 54]. All in all, the case of discrete random vectors seems indeed misaligned with the very essence of the whole copula methodology.

A telling example is the following. Let $X \sim \mathrm{Bern}(\pi_X)$, $Y \sim \mathrm{Bern}(\pi_Y)$ for two probabilities $\pi_X, \pi_Y \in (0, 1)$, and $X \perp\!\!\!\perp Y$ (independent). Then, for reconstructing the bivariate Bernoulli F_{XY} it is enough to plug in (1.1) any copula C such that

$$C(1 - \pi_X, 1 - \pi_Y) = (1 - \pi_X)(1 - \pi_Y), \tag{2.1}$$

as it is seen by inspection. Indeed Sklar's theorem states that C is only identifiable on $\mathrm{Ran}\, F_X \times \mathrm{Ran}\, F_Y = \{0, 1 - \pi_X, 1\} \times \{0, 1 - \pi_Y, 1\}$, but given that C is fixed by trivial constraints along the sides of $\mathbb{I}$, only what happens at $(1 - \pi_X, 1 - \pi_Y)$ may reflect (in)dependence. The product copula (1.2) naturally fulfils (2.1), but so does a wide spectrum of other copulas of miscellaneous shapes whose only common trait is to go through $\left(1 - \pi_X, 1 - \pi_Y, (1 - \pi_X)(1 - \pi_Y)\right) \in (0, 1)^3$ – see Appendix for some illustration. Any conclusion drawn from such a copula-based bivariate Bernoulli model is highly questionable, as its central element C may interchangeably characterise independence or dependence of drastically different strength and nature, yet compatible with (2.1).

It is true that, if one takes two discrete variables X and Y and binds them together through a copula C that we have picked, the '*Conversely*'-part of Sklar's theorem guarantees that one has built a valid bivariate discrete distribution F_{XY} with the right marginals. But there is no special link between C and F_{XY}. For instance, [12] explains how the bivariate Bernoulli distribution on which [18, Example 13] fitted a FGM copula could have been obtained all the same from a Plackett or an Ali-Mikhail-Haq copula, or from the reader's '*peculiar favourite copula family*' [12, p. 128], making it futile to mention any specific copula model at all in this case.

3 Transformations of the margins to uniforms

The root of all trouble is that it is not true that $F_X(X) \sim \mathcal{U}_{[0,1]}$ for X discrete. Though, the $\mathcal{U}_{[0,1]}$-distribution of $F_X(X)$ and $F_Y(Y)$ in the continuous case is clearly what prompted Definition 1.1, and the widespread belief that copula methods are based on transformations of the margins into uniforms. Thus the very foundation of copula theory seems unfit for the discrete framework. Nonetheless, in order to make the discrete case forcibly fit into the classical copula framework, a common practice has been to 'jitter' the original discrete variables with some uniform random noise. The so-created artificial continuous random vector has a unique copula, known as the *checkerboard* copula $C^{\boxplus}$. Arguably, $C^{\boxplus}$ retains some of the dependence structure of the original discrete vector [6, 18, 38, 45, 46], and is a valid copula extension of the underlying subcopula [11, 19, 20]. However, $C^{\boxplus}$ is just a particular choice – and not always the most natural one – among all the copulas satisfying (1.1), and by itself does not solve any of the problems exposed above [45, Remark 1.5(a)].

[35, Section 4] explicitly asked: '*Why does one transform the marginals to a uniform distribution?*', and failing to find any compelling mathematical answer (among other things) lead him to reject the idea of copulas altogether. Yet, it has been widely acknowledged since then [10], but even long before [24, p. 69], that the choice of transforming the margins to uniforms is driven by convenience only. Now, given that transforming to uniform is precisely the stumbling block of copula methods for discrete variables, one may sensibly ask:

why stick to an inessential choice initially made for convenience only, if it is no more convenient at all in the situation of interest?

Sklar's theorem establishes that the valuable information for understanding the joint behaviour of (X, Y) – in particular their dependence – can be captured by some incidental copula C evaluated on Ran $F_X \times$ Ran F_Y. A naive interpretation of this puts C in the foreground, urging us to 'guess' what it might be; whereas it is actually of no importance, making the value of such guesswork unclear. In fact, there is no reason to extend the unique subcopula of a discrete vector to a copula, and any justifiable analysis of the underlying dependence structure should be undertaken at the subcopula level, or equivalent.

Consider again the bivariate Bernoulli case. What Sklar's theorem fundamentally says is that the whole dependence structure can be described by one single number; see the lines following (2.1). Naturally this agrees with any basic analysis of a (2×2)-contingency table, where it is well-known that the dependence/association in the table is captured by a single degree of freedom – cf. the χ^2-test. It is not clear what would be the benefit of playing on a whole bivariate function C for modelling or characterising that dependence, knowing that one single number defines it entirely. [8] argued that this number should be the odds-ratio

$$\omega = \frac{\mathbb{P}(X = 0, Y = 0)\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1, Y = 0)\mathbb{P}(X = 0, Y = 1)} \tag{3.1}$$

because it is 'margin-free' (he did not use that term, though, but see his Corollary 2). [41] calls this 'variation-independence' between the marginal parameters (π_X, π_Y) and the odds ratio ω, and further shows (his Theorem 6.3) that any such margin-free dependence parameter must be a one-to-one function of ω. In Section 5, it will indeed be shown that there is a one-to-one correspondence between the odds ratio and the value $C(1 - \pi_X, 1 - \pi_Y)$ singled out by Sklar's theorem in this situation. It is also easy (Section 5.6) to show that, given π_X and π_Y, the full bivariate pmf can be reconstructed from the value of ω only. Hence the marginal distributions *coupled* with the margin-free dependence parameter ω unequivocally defines the bivariate distribution of interest. Clearly, the single number ω entirely fulfils what we would like the role of a copula to be, while by no means being related to Definition 1.1.

Transformation to uniform marginals is thus clearly not a necessary step for making sense of the main ideas behind copula modelling. Indeed, in Section 4, an alternative perspective on copulas is given, not relying explicitly on PIT. Avoiding PIT allows the concept to be readily adapted to the discrete case as well, while keeping all the pleasant properties of usual copula modelling.

4 Copulas as equivalence classes of dependence

Let (X, Y) be a continuous vector with distribution F_{XY}. For simplicity, assume that X and Y are both supported on $[0, 1]$ (without loss of generality, one can imagine that we observe $X \in \mathbb{R}$ and $Y \in \mathbb{R}$ on the inverse logit scale, and copulas are invariant to monotonic transformations of the margins in any case) and that F_{XY} admits a density f_{XY} with marginal densities f_X and f_Y on the unit square $\mathbb{J}$. Let $\mathcal{F} = \{f : \mathbb{J} \to \mathbb{R}, \text{ s.t. } f \geq 0, \iint_{\mathbb{J}} f = 1\}$, the set of all bivariate probability densities on $\mathbb{J}$, and $\mathbb{S}$ the set of all differentiable strictly increasing functions from $[0, 1]$ to $[0, 1]$. See that $(\mathbb{S}, \circ)$, where $\circ$ denotes function composition, is a group, and so is $(\mathbb{S} \times \mathbb{S}, \circ.)$, where $\circ.$ denotes componentwise composition: for $(\Phi_1, \Psi_1), (\Phi_2, \Psi_2) \in \mathbb{S} \times \mathbb{S}$, $(\Phi_1, \Psi_1) \circ. (\Phi_2, \Psi_2) = (\Phi_1 \circ \Phi_2, \Psi_1 \circ \Psi_2)$.

For any $(\Phi, \Psi) \in \mathbb{S} \times \mathbb{S}$, define $g_{\Phi, \Psi} : \mathcal{F} \to \mathcal{F}$ as

$$g_{\Phi, \Psi}(f)(u, v) = \frac{f(\Phi^{-1}(u), \Psi^{-1}(v))}{\Phi'(\Phi^{-1}(u)) \Psi'(\Psi^{-1}(v))}, \qquad (u, v) \in \mathbb{J}. \tag{4.1}$$

Now, for $(\Phi_1, \Psi_1), (\Phi_2, \Psi_2) \in \mathbb{S} \times \mathbb{S}$, we have that $(g_{\Phi_2, \Psi_2} \circ g_{\Phi_1, \Psi_1})(f) = g_{\Phi_2 \circ \Phi_1, \Psi_2 \circ \Psi_1}(f)$, while if $\Phi(x) = x$ and $\Psi(y) = y$, then $g_{\Phi, \Psi}(f) = f$. This makes $g_{\Phi, \Psi}$ a group action [9, Section 10.1] of $(\mathbb{S} \times \mathbb{S}, \circ.)$ on $\mathcal{F}$, which defines

orbits: for any $f \in \mathcal{F}$, let $[f] = \{f^\star \in \mathcal{F} : \exists (\Phi, \Psi) \in \mathcal{S} \times \mathcal{S} \text{ s.t. } f^\star = g_{\Phi,\Psi}(f)\}$. Such orbits induce an equivalence relation: $f_1 \sim f_2$ whenever $[f_1] = [f_2]$. The quotient space $\overline{\mathcal{F}} = \mathcal{F}/(\mathcal{S} \times \mathcal{S}, \circ.)$ is the set of all equivalence classes.

Clearly $g_{\Phi,\Psi}(f_{XY})$ is the joint density of $(\Phi(X), \Psi(Y))$, i.e., the version of f_{XY} whose marginal distributions have been individually distorted by Φ and Ψ. The class $[f_{XY}]$ contains all those 'marginally distorted' densities. What these densities have in common, compared to other classes, can then only be the constituent of f_{XY} 'between' the margins, that is, what 'glues' them together. This is exactly the definition of 'dependence': '*the information on the law of a random vector which remains to be determined once the marginal laws have been specified*' [52]. Each equivalence class in $\overline{\mathcal{F}}$ is thus representative of a certain dependence structure. The elements of $\overline{\mathcal{F}}$ are really those which deserve the name 'copula', as they genuinely *are* the links ('*copulae*' in Latin) which cement marginals inside bivariate densities. To avoid confusion with Definition 1.1, though, we will call the element $[f] \in \overline{\mathcal{F}}$ the 'nucleus' of f to align with [36]; see Section 1. A nucleus $[f]$ is, in some sense, a bivariate density which has been entirely stripped from its marginals.

Now, the abstract concept of a bivariate density with no marginals is difficult to visualise. Hence, for understanding the inner dependence structure of the vector (X, Y), one may want to exhibit a simple re-embodiment of $[f_{XY}]$ into a proper density by pasting on it some default marginals. If one targets uniform marginals, then, by PIT, this is the element $g_{\Phi,\Psi}(f_{XY}) \in [f_{XY}]$ which corresponds to $(\Phi, \Psi) = (F_X, F_Y)$. That particular representative is thus

$$\bar{f}_{XY}(u, v) = \frac{f_{XY}(F_X^{-1}(u), F_Y^{-1}(v))}{f_X(F_X^{-1}(u))f_Y(F_Y^{-1}(v))} \in [f_{XY}],$$

in which we recognise the density c of the copula C of F_{XY}. Clearly the choice of uniform margins for re-embodying $[f_{XY}]$ into a proper density is totally arbitrary. It seems just sensible, for interpretation and visualisation purpose, to keep things as uncomplicated as possible. That said, uniforms and/or PIT do not play any role when defining the concept of nucleus, which is really what copulas are all about. The construction can thus be adapted *mutatis mutandis* to discrete distributions, as detailed below.

5 The Bernoulli copula

5.1 The bivariate Bernoulli distribution

Consider again the case of two Bernoulli random variables $X \sim \text{Bern}(\pi_X)$ and $Y \sim \text{Bern}(\pi_Y)$ sharing (potentially) some dependence. The corresponding bivariate Bernoulli distribution, say **p**, is typically presented under the form of a (2×2)-table:

$x \backslash Y$	0	1	
0	p_{00}	p_{01}	$p_{0\bullet}$
1	p_{10}	p_{11}	$p_{1\bullet}$
	$p_{\bullet 0}$	$p_{\bullet 1}$	1

$$\tag{5.1}$$

where for $x, y \in \{0, 1\}$, $p_{xy} = \mathbb{P}(X = x, Y = y)$, $p_{\bullet y} = p_{0y} + p_{1y}$ and $p_{x\bullet} = p_{x0} + p_{x1}$, with $\pi_X = p_{1\bullet}$ and $\pi_Y = p_{\bullet 1}$. Assume $0 < \pi_X, \pi_Y < 1$ (non-degenerate table). Define $\mathcal{P}_{2\times 2}$ the set of all such bivariate Bernoulli probability mass functions, where each $\mathbf{p} \in \mathcal{P}_{2\times 2}$ is identified to the matrix

$$\mathbf{p} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \tag{5.2}$$

Now, as $\sum_{x,y} p_{xy} = 1$, one can actually identify $\mathcal{P}_{2\times 2}$ to the 3-dimensional simplex, here a regular tetrahedron whose vertices are the degenerate distributions $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{d}_3$ and $\mathbf{d}_4$; see Figure 5.1. We will call this tetrahedron the *Bernoulli tetrahedron*. Note that, as we assume $0 < \pi_X, \pi_Y < 1$, the 4 vertices and the edges $\mathbf{d}_1\mathbf{d}_2$, $\mathbf{d}_1\mathbf{d}_3$, $\mathbf{d}_2\mathbf{d}_4$ and $\mathbf{d}_3\mathbf{d}_4$ are not admissible elements of $\mathcal{P}_{2\times 2}$.
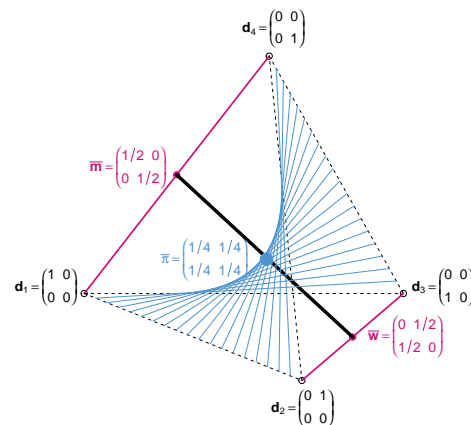
Figure 5.1: The Bernoulli tetrahedron. The blue surface is the Bernoulli nucleus of independence. The thick line $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$ is the locus of all the Bernoulli copula pmf's. The centre of the tetrahedron is the Bernoulli independence copula $\overline{\pi}$. The magenta edges represent 'absolute association' (positive, edge $\mathbf{d}_1\mathbf{d}_4$; negative, edge $\mathbf{d}_2\mathbf{d}_3$). Their mid-points $\overline{\mathbf{m}}$ and $\overline{\mathbf{w}}$ are the upper and lower Fréchet bounds for Bernoulli copulas.

5.2 Marginal transformations

Mimicking Section 4, we look for isolating the constituent of $\mathbf{p}$ which remains invariant to 'monotonic distortion' of the margins. In the continuous case, by such distortion it was meant the vector $(\Phi(X), \Psi(Y))$ whose density (4.1) can take miscellaneous shapes. Here, this 'transformation trick' does not work: for $X \sim \text{Bern}(\pi_X)$, $\Phi(X)$ remains the same two-point distribution $(1-\pi_X, \pi_X)$ (only the 'labels' change), and same for $\Psi(Y)$. Now, one can see (4.1) from a more basic perspective, considering the group action $g_{\Phi,\Psi}$ as just a mechanism re-assigning the initial probability mass differently over $\mathbb{I}$. Under the effect of $g_{\Phi,\Psi}$, the point (u, v), initially assigned the probability $f_{XY}(u, v)\, du\, dv$, would now get a probability $f_{XY}^{\star}(u, v)\, du\, dv$ where $f_{XY}^{\star}(u, v)$ is given by (4.1) – in a sense, we regard this as the effect of a transport map taking the initial probability measure onto another one, as opposed to the initial interpretation as the resulting probability measure after transformation of the random variables. Note that the normalisation by $\Phi'(\Phi^{-1}(u))\,\Psi'(\Psi^{-1}(v))$ just guarantees $\int f_{XY}^{\star} = 1$. This more fundamental interpretation of (4.1) carries over to the Bernoulli framework.

Indeed, for some $\phi > 0$, define a distorted distribution for X as $\text{Bern}(\pi_X^{\star})$, where $\pi_X^{\star} = \frac{\phi\pi_X}{1-\pi_X+\phi\pi_X}$. Clearly, for $\phi > 1$, $\pi_X^{\star} > \pi_X$: some of the probability initially assigned to $X = 0$ has been transferred to the next value, $X = 1$; and reversely for $\phi < 1$. Like above, the factor $1/(1 - \pi_X + \phi\pi_X)$ is just a normalisation, guaranteeing $\pi_X^{\star} \in [0, 1]$ for all $\phi > 0$. When the margin Y is similarly distorted, the initial joint probability distribution is re-assigned through table (5.1) by a similar process of transferring probability weight between adjacent cells. The organisation of the cells, in particular their order along each margin, is not altered: the marginal distortions are monotonic in that sense. In effect, the distorted table is obtained by multiplying the rows and columns of (5.2) by positive values (and renormalise). This totally concords with what [53] and [36] urged; see Section 1.

Specifically, define $\mathcal{D}_{2\times2}^{(1)}$ the set of all diagonal matrices whose entry $(1, 1)$ is equal to 1, and for any $\phi, \psi > 0$, set

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix} \in \mathcal{D}_{2\times2}^{(1)} \quad \text{and} \quad \Psi = \begin{pmatrix} 1 & 0 \\ 0 & \psi \end{pmatrix} \in \mathcal{D}_{2\times2}^{(1)}.$$

Equipped with the matrix multiplication $\cdot$, $(\mathcal{D}_{2\times2}^{(1)}, \cdot)$ is a group, and so is $(\mathcal{D}_{2\times2}^{(1)}\times\mathcal{D}_{2\times2}^{(1)}, \cdot.)$, where $\cdot.$ is componen-twise matrix multiplication: for $(\Phi_1, \Psi_1), (\Phi_2, \Psi_2) \in (\mathcal{D}_{2\times2}^{(1)}\times\mathcal{D}_{2\times2}^{(1)})$, $(\Phi_1, \Psi_1) \cdot. (\Phi_2, \Psi_2) = (\Phi_1 \cdot \Phi_2, \Psi_1 \cdot \Psi_2)$.

Now define $g_{\Phi, \Psi} : \mathcal{P}_{2\times 2} \to \mathcal{P}_{2\times 2}$:

$$g_{\Phi, \Psi}(\mathbf{p}) = \frac{\Phi \cdot \mathbf{p} \cdot \Psi}{\|\Phi \cdot \mathbf{p} \cdot \Psi\|_1} = \frac{1}{p_{00} + \psi p_{01} + \phi p_{10} + \phi \psi p_{11}} \begin{pmatrix} p_{00} & \psi p_{01} \\ \phi p_{10} & \phi \psi p_{11} \end{pmatrix}. \tag{5.3}$$

Similarly to Section 4, $g_{\Phi, \Psi}$ is a group action of $(\mathcal{D}_{2\times 2}^{(1)} \times \mathcal{D}_{2\times 2}^{(1)}, \cdot)$ on $\mathcal{P}_{2\times 2}$. Any $\mathbf{p} \in \mathcal{P}_{2\times 2}$ induces an orbit $[\mathbf{p}] = \{\mathbf{p}^\star \in \mathcal{P}_{2\times 2} : \exists (\Phi, \Psi) \in \mathcal{D}_{2\times 2}^{(1)} \times \mathcal{D}_{2\times 2}^{(1)} \text{ s.t. } \mathbf{p}^\star = g_{\Phi, \Psi}(\mathbf{p})\}$. Those orbits form equivalence classes of bivariate Bernoulli distributions: $\mathbf{p}_1 \sim \mathbf{p}_2$ whenever $[\mathbf{p}_1] = [\mathbf{p}_2]$. The quotient space $\overline{\mathcal{P}}_{2\times 2} = \mathcal{P}_{2\times 2}/(\mathcal{D}_{2\times 2}^{(1)} \times \mathcal{D}_{2\times 2}^{(1)}, \cdot)$ is the set of all those equivalence classes. Any $[\mathbf{p}] \in \overline{\mathcal{P}}_{2\times 2}$ is a class of bivariate Bernoulli distributions (5.2) sharing the same 'core' structure once we strip them from their margins.

5.3 Bernoulli nucleus and Bernoulli copula probability mass function

For any distribution $\mathbf{p} \in \mathcal{P}_{2\times 2}$, define

$$\omega : \mathcal{P}_{2\times 2} \to \mathbb{R}^+ : \omega(\mathbf{p}) = \frac{p_{00} p_{11}}{p_{10} p_{01}},$$

the odds-ratio (3.1). It is obvious that, for any $(\Phi, \Psi) \in \mathcal{D}_{2\times 2}^{(1)} \times \mathcal{D}_{2\times 2}^{(1)}$,

$$\omega(g_{\Phi, \Psi}(\mathbf{p})) = \frac{p_{00} \phi \psi p_{11}}{\phi p_{10} \psi p_{01}} = \frac{p_{00} p_{11}}{p_{10} p_{01}} = \omega(\mathbf{p}), \tag{5.4}$$

establishing that the elements of $\overline{\mathcal{P}}_{2\times 2}$ are again classes of equivalent dependence. We call $[\mathbf{p}]$ the nucleus of $\mathbf{p}$, which contains the full information about how the Bernoulli marginals are glued together inside $\mathbf{p}$.

In the Bernoulli tetrahedron, the sets of distributions $\mathbf{p} \in \mathcal{P}_{2\times 2}$ sharing common odds ratios $\omega \in (0, \infty)$ are doubly-ruled surfaces corresponding to sections of hyperboloids of one sheet [15, Section 3]. E.g., Figure 5.1 shows the surface corresponding to $\omega = 1$, that is, all bivariate Bernoulli distributions for which $X \perp\!\!\!\perp Y$. Running ω over $(0, \infty)$ produces similar non-intersecting surfaces, which are the Bernoulli nuclei $[\mathbf{p}] \in \overline{\mathcal{P}}_{2\times 2}$. For the cases $\omega \in \{0, \infty\}$, see Section 5.4.

For interpretation purpose, it may be insightful to define a representative of $[\mathbf{p}]$, that is, a particular 'simple' bivariate Bernoulli distribution with odds-ratio $\omega(\mathbf{p})$. Again, a natural choice is the element of $[\mathbf{p}]$ with uniform margins, as [21] suggested (Section 1). Simple algebra reveals that, for $\mathbf{p} \in \mathcal{P}_{2\times 2}$ such that $\omega(\mathbf{p}) = \omega \geq 0$, there is a unique element in $[\mathbf{p}]$ with Bernoulli$(1/2)$-margins, which is

$$\overline{\mathbf{p}} = \begin{pmatrix} \frac{\sqrt{\omega}}{2(1+\sqrt{\omega})} & \frac{1}{2(1+\sqrt{\omega})} \\ \frac{1}{2(1+\sqrt{\omega})} & \frac{\sqrt{\omega}}{2(1+\sqrt{\omega})} \end{pmatrix}. \tag{5.5}$$

This representative is akin to the copula density in the continuous case, hence we call $\overline{\mathbf{p}}$ the *Bernoulli copula probability mass function* (copula pmf). Note that the values in (5.5) were mentioned in [4, 11.2-14], while a similar 'copula' was investigated in [51].

In the Bernoulli tetrahedron, all $\mathbf{p} \in \mathcal{P}_{2\times 2}$ with the same marginal distributions must lie on a straight line orthogonal to the edges $\mathbf{d}_1 \mathbf{d}_4$ and $\mathbf{d}_2 \mathbf{d}_3$ [15, Section 4]. Denote

$$\overline{\mathbf{m}} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \qquad \text{and} \qquad \overline{\mathbf{w}} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \tag{5.6}$$

the mid-points of $\mathbf{d}_1 \mathbf{d}_4$ and $\mathbf{d}_2 \mathbf{d}_3$, respectively. The segment $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$, shown in Figure 5.1, is orthogonal to both $\mathbf{d}_1 \mathbf{d}_4$ and $\mathbf{d}_2 \mathbf{d}_3$. Hence it consists of all those distributions which share the same margins as $\overline{\mathbf{m}}$ and $\overline{\mathbf{w}}$, which are Bernoulli$(1/2)$ for both X and Y, that is, all the Bernoulli copula pmf's. The element (5.5) is the unique intersection between $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$ and the nucleus $[\mathbf{p}]$ characterised by the odds ratio $\omega(\mathbf{p}) = \omega$.

Given that $X \perp\!\!\!\perp Y \iff \omega = 1$, the independence Bernoulli copula pmf is evidently

$$\overline{\pi} \doteq \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix},$$

and clearly $X \perp\!\!\!\perp Y \iff \overline{\mathbf{p}} = \overline{\pi}$. This can be contrasted to the observation made in Section 2 that a continuous copula C gluing two independent Bernoulli's as in (1.1) need not be the independence copula.

Example 5.1. We call '*confetti plot*' the below – somewhat naive but effective – visual display of bivariate Bernoulli pmf's and their copulas. The size and the colour of the dots are proportional to the value of the corresponding probability. Marginal probabilities are shown as black dots on the same scale. Figure 5.2 shows such plots for [53]'s comparison of three hospitals on vaccination and recovery for smallpox patients. The top row shows the initial bivariate Bernoulli distributions as tables like (5.1) (left: Sheffield; middle: Leicester; right: Homerton and Fulham) – details and exact figures to be found in [53, Tables I, III, IV]. The nature of the dependence and how it compares across those three distributions is not obvious visually, owing to their different and largely unbalanced margins. The bottom row shows the corresponding Bernoulli copula pmf's, together with the respective values of ω (and Υ, see Section 5.5). Those copula pmf's make it clear, visually, that the association is positive and of similar magnitude across the three distributions (although with a slight decrease from left to right). □

ω = 19.4995, Y = 0.6307   ω = 13.462, Y = 0.5716   ω = 9.18, Y = 0.5037
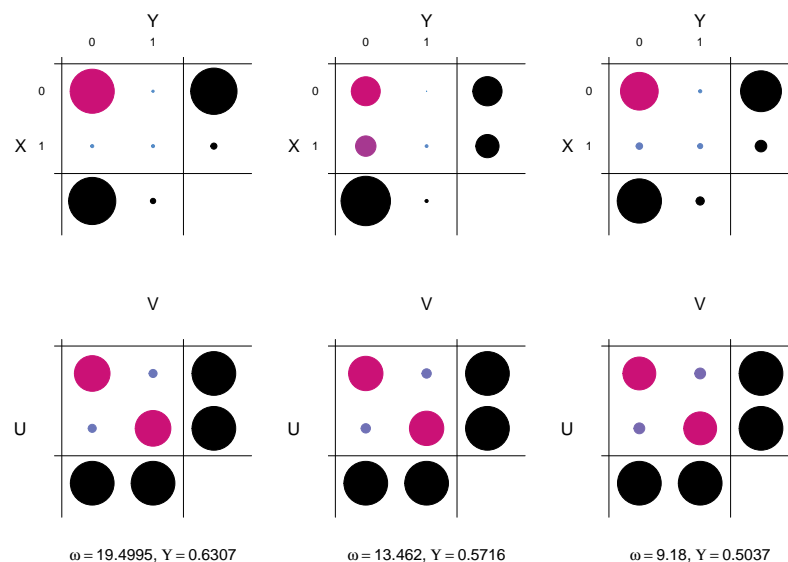
Figure 5.2: Confetti plots of Bernoulli pmf's (top row) and Bernoulli copula pmf's (bottom row) for [53]'s smallpox vaccination data: left: Sheffield; middle: Leicester; right: Homerton and Fulham.

5.4 Structural zeros

The limit values $\omega = 0$ and $\omega = \infty$ occur when (at least) one of the entries of (5.2) is 0, what is usually referred to as a 'structural zero' of $\mathbf{p}$ [4, Chapter 5]. Here, for $\omega = 0$ or $\omega = \infty$, we respectively get from (5.5) the copula pmf's $\overline{\mathbf{w}}$ and $\overline{\mathbf{m}}$ (5.6). It can be checked that $\overline{\mathbf{w}}$ and $\overline{\mathbf{m}}$ are the pmf's associated to the Fréchet lower and upper bounds [27, Section 2.9] in the class of bivariate Bernoulli distributions with Bern$(1/2)$-margins, in agreement with the values $\omega = 0$ and $\omega = \infty$ indicating maximal negative and positive association between X and Y. In the Bernoulli tetrahedron (Figure 5.1), any Bernoulli copula pmf lies between the 'bounds' $\overline{\mathbf{w}}$ and $\overline{\mathbf{m}}$ on the segment $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$, indeed.

The value $\omega = 0$ arises in distributions like

$$(i) \quad \mathbf{p}_1 = \begin{pmatrix} 0 & \times \\ \times & 0 \end{pmatrix}; \quad \text{or} \quad (ii) \quad \mathbf{p}_2 = \begin{pmatrix} 0 & \times \\ \times & \times \end{pmatrix}, \mathbf{p}_3 = \begin{pmatrix} \times & \times \\ \times & 0 \end{pmatrix}, \quad (5.7)$$

where ×'s represent non-zero elements. In the terminology of [29, Sections 33.6-33.7], case (i) corresponds to *absolute* association, whereas cases (ii) correspond to *complete* association. Clearly (i) represents 'perfect dependence' (negative, in this case), but it is not that clear for (ii) as there is no one-to-one correspondence between X and Y. Therefore, the fact that the odds-ratio $\omega = 0$ and the corresponding copula pmf $\overline{\mathbf{w}}$ do not distinguish between (i) and (ii) may appear puzzling. Yet it is easily seen that 'absolute association' (i) is only possible if $\pi_X + \pi_Y = 1$. Whenever $\pi_X + \pi_Y \neq 1$, any sense of 'perfect dependence' automatically translates into 'complete association'. Hence the dependence is actually as strong as can be in both cases (i) and (ii) *given the margins*. A marginal feature, the distinction between (i) and (ii) must be ignored by the copula pmf.

In the Bernoulli tetrahedron, distributions showing 'absolute association' ($\mathbf{p}_1$) lie on the edge $\mathbf{d}_2\mathbf{d}_3$, while 'complete association' amounts to the two (open) faces $\mathbf{d}_2\mathbf{d}_3\mathbf{d}_4$ (distributions of type $\mathbf{p}_2$) and $\mathbf{d}_1\mathbf{d}_2\mathbf{d}_3$ (distributions of type $\mathbf{p}_3$). Actually, the union $(\mathbf{d}_2\mathbf{d}_3 \cup \mathbf{d}_2\mathbf{d}_3\mathbf{d}_4 \cup \mathbf{d}_1\mathbf{d}_2\mathbf{d}_3)$ forms the limit of the surfaces of constant odds ratio ω as $\omega \to 0$. However, there is no transformation (5.3) making $\mathbf{p}_2$ or $\mathbf{p}_3$ into $\mathbf{p}_1$, or even $\mathbf{p}_2$ into $\mathbf{p}_3$, as a new zero would need to be created, which is not possible given that ϕ and ψ are positive in (5.3). Clearly, $[\mathbf{p}_1] = \mathbf{d}_2\mathbf{d}_3$, $[\mathbf{p}_2] = \mathbf{d}_2\mathbf{d}_3\mathbf{d}_4$ and $[\mathbf{p}_3] = \mathbf{d}_1\mathbf{d}_2\mathbf{d}_3$ are distinct elements of $\overline{\mathcal{P}}_{2\times2}$.

The edge $[\mathbf{p}_1] = \mathbf{d}_2\mathbf{d}_3$ physically intersects with $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$ (at $\overline{\mathbf{w}}$, obviously), thus $\overline{\mathbf{w}} \in [\mathbf{p}_1]$. By contrast, the (open) faces $[\mathbf{p}_2] = \mathbf{d}_2\mathbf{d}_3\mathbf{d}_4$ and $[\mathbf{p}_3] = \mathbf{d}_1\mathbf{d}_2\mathbf{d}_3$ do not intersect with $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$, thus $\overline{\mathbf{w}} \notin [\mathbf{p}_2] \cup [\mathbf{p}_3]$. What is true, though, is that $\overline{\mathbf{w}}$ is a limit point of $[\mathbf{p}_2]$ and $[\mathbf{p}_3]$, as one can approach $\overline{\mathbf{w}}$ arbitrarily close while staying on any of the two faces. In these critical cases (ii), it is thus necessary to extend the nuclei $[\mathbf{p}_2]$ or $[\mathbf{p}_3]$ to their closure for them to include the corresponding copula pmf $\overline{\mathbf{w}}$. Further characterisation will be given in more generality in Section 6. The case $\omega = \infty$ and $\overline{\mathbf{p}} = \overline{\mathbf{m}}$ is treated in perfect analogy.

Remark 5.1. Although $\omega(\mathbf{p}_1) = \omega(\mathbf{p}_2) = \omega(\mathbf{p}_3) = 0$, $[\mathbf{p}_1]$, $[\mathbf{p}_2]$ and $[\mathbf{p}_3]$ are distinct, hence there is no one-to-one correspondence between $[\mathbf{p}]$ and $\omega(\mathbf{p})$. What differentiates $[\mathbf{p}_1]$, $[\mathbf{p}_2]$ and $[\mathbf{p}_3]$ is the different layout of structural zeros of their elements, that is, their support $\text{Supp}(\mathbf{p}) = \{(x, y) \in \{0, 1\} \times \{0, 1\} : p_{xy} > 0\}$. In fact, any nucleus $[\mathbf{p}]$ is unequivocally characterised by the couple $(\text{Supp}(\mathbf{p}), \omega(\mathbf{p}))$. This double characterisation clarifies what really makes the dependence structure of a bivariate Bernoulli vector (X, Y). The dominant effect is the presence of structural zeros and their layout. In fact, any such zero in $\mathbf{p}$ implies, by definition, the incompatibility of two particular values taken by X and Y. Hence it is a critical constituent of the dependence between X and Y, to the point of making the value of the other non-null p_{xy}'s irrelevant. When all p_{xy}'s are positive, then the dependence is more subtle, and characterised by ω.

5.5 Yule's colligation coefficient

Suppose that (U, V) is a bivariate Bernoulli vector with joint pmf (5.5) for some $\omega \geq 0$. One can check that Pearson's correlation between U and V is

$$\Upsilon = \frac{\sqrt{\omega} - 1}{\sqrt{\omega} + 1}, \quad (5.8)$$

which is exactly Yule's '*colligation coefficient*' [53, pp. 592-593]. Hence Υ can be regarded as the Bernoulli analogue to Spearman's ρ in the sense that it is Pearson's correlation computed *after copula transformation*. As a one-to-one function of ω, Υ is margin-free. Besides, $\Upsilon = 0 \iff \omega = 1 \iff X \perp\!\!\!\perp Y$ in (5.1), and $|\Upsilon|$ attains its maximum value 1 when $\omega = 0$ or $\omega = \infty$, which corresponds to the Fréchet bounds (5.6). This can be contrasted to the inconsistencies observed for the usual Spearman's ρ computed on discrete vectors [18, Sections 4.2 and 4.4].

In fact, as $\sqrt{\omega} = (1 + \Upsilon)/(1 - \Upsilon)$, the copula pmf $\overline{\mathbf{p}}$ (5.5) can be written under the even simpler form

$$\overline{\mathbf{p}} = \frac{1}{4} \begin{pmatrix} 1 + \Upsilon & 1 - \Upsilon \\ 1 - \Upsilon & 1 + \Upsilon \end{pmatrix}$$

in terms of $\Upsilon \in [-1, 1]$. The effect of Υ on $\overline{\mathbf{p}}$ is thus linear in nature. The value of Υ acts as a ruler along $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$ in Figure 5.1: from $\Upsilon = -1$ at $\overline{\mathbf{w}}$ to $\Upsilon = 1$ at $\overline{\mathbf{m}}$, via $\Upsilon = 0$ at $\overline{\boldsymbol{\pi}}$.

Kendall's τ corrected for the occurrence of ties, i.e. 'τ_b' [18, Definition 3], is, for the bivariate Bernoulli case, $\tau_b = \frac{p_{00} - p_{0\bullet} p_{\bullet 0}}{\sqrt{p_{0\bullet} p_{\bullet 0} p_{1\bullet} p_{\bullet 1}}}$. This, computed on the copula pmf (5.5), reduces down to $\tau_b = \Upsilon$ again. The above observations suggest that Yule's Υ is a very natural, if not the canonical, dependence parameter in the (2×2)-table framework. This is noteworthy as, although it was originally Yule's preferred association measure [53, p. 592], this coefficient has by now largely passed into oblivion.

5.6 Construction of arbitrary bivariate Bernoulli distributions with given copula pmf

Analogously to the '*Conversely*'-part of Theorem 1.1, one can wonder if it is always possible to construct a bivariate Bernoulli distribution $\mathbf{p}$ whose marginals are $\mathrm{Bern}(\pi_X)$ and $\mathrm{Bern}(\pi_Y)$ ($0 < \pi_X, \pi_Y < 1$) and dependence structure prescribed by a certain Bernoulli copula (5.5). The answer is affirmative. In Figure 5.1, the set of all bivariate Bernoulli's with the requested margins is a segment parallel to $\overline{\mathbf{w}}\,\overline{\mathbf{m}}$ [14, Section 4], which, for $0 < \omega < \infty$, necessarily intersects uniquely with the surface/nucleus $[\mathbf{p}]$ characterised by $\omega(\mathbf{p}) = \omega$. Following [36, equation (2$\star$)], p_{11} in $\mathbf{p}$ is, for $\omega \neq 1$,

$$p_{11} = \frac{1}{2(\omega - 1)} \left\{ 1 + (\omega - 1)(\pi_X + \pi_Y) - \sqrt{[1 + (\omega - 1)(\pi_X + \pi_Y)]^2 - 4\omega(\omega - 1)\pi_X\pi_Y} \right\}. \tag{5.9}$$

(If $\omega = 1$, then trivially $p_{11} = \pi_X\pi_Y$.) The other values follow by substitution. In particular, $p_{00} = 1 - \pi_X - \pi_Y + p_{11}$. As p_{11} in (5.9) is an increasing function of ω, so is p_{00}. In the same time, 'the value $C(1 - \pi_X, 1 - \pi_Y)$ singled out by Sklar's theorem', described below (3.1), is precisely p_{00}, establishing the one-to-one correspondence between $C(1 - \pi_X, 1 - \pi_Y)$ and ω. Characterising the dependence in a bivariate Bernoulli vector by ω, or any monotonic function thereof, is thus totally consistent with Sklar's theorem.

For $\omega = 0$, we must have either $p_{00} = 0$ or $p_{11} = 0$ (or both). By obvious substitution, one gets

$$\mathbf{p}_1 = \begin{pmatrix} 0 & 1 - \pi_X \\ 1 - \pi_Y & 0 \end{pmatrix}, \quad \mathbf{p}_2 = \begin{pmatrix} 0 & 1 - \pi_X \\ 1 - \pi_Y & \pi_X + \pi_Y - 1 \end{pmatrix} \quad \text{or} \quad \mathbf{p}_3 = \begin{pmatrix} 1 - \pi_X - \pi_Y & \pi_Y \\ \pi_X & 0 \end{pmatrix}, \tag{5.10}$$

if $\pi_X + \pi_Y = 1$, $\pi_X + \pi_Y > 1$ or $\pi_X + \pi_Y < 1$, respectively. These distributions correspond to the lower Fréchet bound in the class of bivariate Bernoulli distributions with the requested margins [27, Example 2.9].

Remark 5.2. In the Bernoulli tetrahedron, these distributions lie at the intersection between the 'segment of requested margins', and either the face $\mathbf{d}_2\mathbf{d}_3\mathbf{d}_4$ or the face $\mathbf{d}_1\mathbf{d}_2\mathbf{d}_3$. Because that segment is parallel to $\overline{\mathbf{w}}\overline{\mathbf{m}}$, it can only pierce one of those two faces, establishing geometrically the uniqueness of a distribution $\mathbf{p}$ with the requested margins and $\omega = 0$. □

The case $\omega = \infty$ is treated identically, by symmetry. Hence any two Bernoulli distributions can be glued together through the Bernoulli copula for showing any level of dependence set by $\omega \in [0, \infty]$. The extension of those ideas to general $(R \times S)$-discrete distributions is investigated in the next section.

6 General discrete distributions with finite support

Let (X, Y) be a bivariate discrete vector where X and Y may only take a finite number of values. Without loss of generality, let $X \in \mathcal{S}_X \doteq \{0, 1, \ldots, R - 1\}$ and $Y \in \mathcal{S}_Y \doteq \{0, 1, \ldots, S - 1\}$, with $R, S \in \mathbb{N}$, $2 \leq R, S < \infty$.

Let $\mathbf{p}$ be its joint probability mass function, defined by $p_{xy} = \mathbb{P}(X = x, Y = y)$, $(x, y) \in \mathcal{S}_X \times \mathcal{S}_Y$, and $\mathbf{p}_X = (p_{0\bullet}, p_{1\bullet}, \ldots, p_{R-1\bullet})$ and $\mathbf{p}_Y = (p_{\bullet 0}, p_{\bullet 1}, \ldots, p_{\bullet S-1})$ its marginal distributions: $p_{x\bullet} = \sum_{y \in \mathcal{S}_Y} p_{xy} = \mathbb{P}(X = x)$ and $p_{\bullet y} = \sum_{x \in \mathcal{S}_X} p_{xy} = \mathbb{P}(Y = y)$. Let $\mathcal{P}_{R \times S}$ be the set of all such bivariate discrete distributions $\mathbf{p}$ with $p_{x\bullet} > 0 \ \forall x \in \mathcal{S}_X$ and $p_{\bullet y} > 0 \ \forall y \in \mathcal{S}_Y$, identified to the $(R \times S)$-matrices

$$\mathbf{p} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0,S-1} \\ p_{10} & p_{11} & \cdots & p_{1,S-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{R-1,0} & p_{R-1,1} & \cdots & p_{R-1,S-1} \end{pmatrix}.$$

Any such distribution can be regarded as a point in the $(RS - 1)$-dimensional simplex [13].

6.1 Odds ratio matrix

As $\mathrm{Ran}\, F_X \times \mathrm{Ran}\, F_Y$ here consists of $(R - 1)(S - 1)$ informative locations (i.e., strictly inside the unit square), Sklar's Theorem implies that one must be able to entirely describe the inner dependence structure of $\mathbf{p}$ by $(R - 1)(S - 1)$ parameters, naturally in agreement with the usual break down of degrees of freedom in comparable $(R \times S)$-contingency tables. Those $(R - 1)(S - 1)$ parameters can be a family of odds-ratios [2]. [41, p. 119] spells out that those odds-ratios are margin-free ('variation-independent' of the marginal parameters). [1, p. 55] stressed that '*given the marginals, the odds ratios determine the cell probabilities*'; in other words, the full distribution can be entirely reconstructed by *coupling* the marginal distributions and the margin-free set of odds ratios. Again, those entirely fulfil here the desired role of classical copulas, the explicit resort to which being purposeless.

Let

$$\omega_{xy} = \frac{p_{00} p_{xy}}{p_{0y} p_{x0}}, \quad \forall (x, y) \in \mathcal{S}_X \backslash \{0\} \times \mathcal{S}_Y \backslash \{0\}. \tag{6.1}$$

Let $\mathcal{M}^{(+)}_{(R-1) \times (S-1)}$ be the set of all $(R - 1) \times (S - 1)$-matrices with non-negative, possibly infinite, entries. Define the map

$$\Omega : \mathcal{P}_{R \times S} \to \mathcal{M}^{(+)}_{(R-1) \times (S-1)} : \mathbf{p} \to \Omega(\mathbf{p}) = [\omega_{xy}]_{\substack{x=1,\ldots,R-1, \\ y=1,\ldots,S-1}} \tag{6.2}$$

where ω_{xy} is given by (6.1). Call $\Omega(\mathbf{p})$ the *odds ratio matrix*.

Remark 6.1. Although they were ruled out in (5.2)-(5.4) when assuming $0 < \pi_X, \pi_Y < 1$, cases of $0/0$ may arise in (6.1). Then the corresponding entry of $\Omega(\mathbf{p})$ may be left undefined. Admitting some slight lack of rigour, we identify two odds ratio matrices whose all well defined entries are equal, i.e., an undefined entry in a matrix is assumed to be equal to whatever the corresponding entry may be in the other. $\square$

6.2 Marginal transformations, nucleus and copula probability mass function

Define $\mathcal{D}^{(1)}_{Q \times Q}$ the set of all diagonal $Q \times Q$ matrices whose entry $(1, 1)$ is equal to 1 and other diagonal entries are positive. Similarly to (5.3), for any $\Phi \in \mathcal{D}^{(1)}_{R \times R}$ and $\Psi \in \mathcal{D}^{(1)}_{S \times S}$, let

$$g_{\Phi,\Psi} : \mathcal{P}_{R \times S} \to \mathcal{P}_{R \times S} : g_{\Phi,\Psi}(\mathbf{p}) = \frac{\Phi \cdot \mathbf{p} \cdot \Psi}{\|\Phi \cdot \mathbf{p} \cdot \Psi\|_1}. \tag{6.3}$$

The matrix Φ multiplies the rows of $\mathbf{p}$ and the matrix Ψ multiplies the columns of $\mathbf{p}$: this is akin to 'marginal distortions' as in Section 5.2. This defines a group action on $\mathcal{P}_{R \times S}$, which induces orbits $[\mathbf{p}]$:

$$\mathbf{p} \sim \mathbf{p}^\star \iff [\mathbf{p}] = [\mathbf{p}^\star] \iff \exists (\Phi, \Psi) \in \mathcal{D}^{(1)}_{R \times R} \times \mathcal{D}^{(1)}_{S \times S} \text{ s.t. } \mathbf{p}^\star = g_{\Phi,\Psi}(\mathbf{p}). \tag{6.4}$$

Further, define a limit point of $[\mathbf{p}]$ as an element of $\mathcal{P}_{R \times S}$ which can be written as

$$\frac{\left(\prod_{k=1}^{\infty} \Phi_k\right) \cdot \mathbf{p} \cdot \left(\prod_{k=1}^{\infty} \Psi_k\right)}{\| \left(\prod_{k=1}^{\infty} \Phi_k\right) \cdot \mathbf{p} \cdot \left(\prod_{k=1}^{\infty} \Psi_k\right) \|_1} \tag{6.5}$$

for some sequences of matrices $\Phi_1, \Phi_2, \ldots \in \mathcal{D}_{R \times R}^{(1)}$ and $\Psi_1, \Psi_2, \ldots \in \mathcal{D}_{S \times S}^{(1)}$. Let $\mathrm{Cl}([\mathbf{p}])$ be the closure of $[\mathbf{p}]$, that is, the union of $[\mathbf{p}]$ and its limit points.

Free from any sense of marginal distributions, the orbits $[\mathbf{p}]$ must again be equivalence classes of dependence. Indeed, for any two $\mathbf{p}, \mathbf{p}^\star \in \mathcal{P}_{R \times S}$, $\mathbf{p} \sim \mathbf{p}^\star \Rightarrow \Omega(\mathbf{p}) = \Omega(\mathbf{p}^\star)$, as all odds ratios are preserved by $g_{\Phi, \Psi}$, exactly as in (5.4). This holds true for any undefined elements of $\Omega(\mathbf{p})$ as well, as $g_{\Phi, \Psi}$ leaves the zeros of $\mathbf{p}$ unaffected. Hence $[\mathbf{p}]$ will again be called the nucleus of the discrete pmf $\mathbf{p}$. If all entries of $\Omega(\mathbf{p})$ are defined and positive, then $\Omega(\mathbf{p}) = \Omega(\mathbf{p}^\star) \Rightarrow [\mathbf{p}] = [\mathbf{p}^\star]$. Like in Remark 5.1, though, one may find two $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}_{R \times S}$ with $\Omega(\mathbf{p}_1) = \Omega(\mathbf{p}_2)$ but $[\mathbf{p}_1] \neq [\mathbf{p}_2]$ when $\mathrm{Supp}(\mathbf{p}_1) \neq \mathrm{Supp}(\mathbf{p}_2)$, that is, when $\mathbf{p}_1$ and $\mathbf{p}_2$ show a different pattern of structural zeros. Again, the preponderant role of structural zeros on the dependence structure appears clearly.

Now one may wish to single out the member of $[\mathbf{p}]$ with uniform marginals for embodying the dependence pattern in $\mathbf{p}$ by a simple element of the class. That one would be called the 'copula pmf' of $\mathbf{p}$, leading to the following definition of a discrete copula, the obvious analogue to Definition 1.1.

Definition 6.1. *A bivariate $(R \times S)$-discrete copula is the bivariate $(R \times S)$-discrete distribution of a vector (U, V) whose both marginal distributions are discrete uniform on $\mathcal{S}_U \doteq \{\frac{1}{R+1}, \frac{2}{R+1}, \ldots, \frac{R}{R+1}\}$ and $\mathcal{S}_V \doteq \{\frac{1}{S+1}, \frac{2}{S+1}, \ldots, \frac{S}{S+1}\}$, respectively. The associated copula probability mass function (copula pmf) is thus a bivariate discrete pmf $\overline{\mathbf{p}}$ on $\mathcal{S}_U \times \mathcal{S}_V$ such that for all $u \in \{0, \ldots, R-1\}$, $\sum_{v=0}^{S-1} \overline{p}_{uv} = \frac{1}{R}$, and for all $v \in \{0, \ldots, S-1\}$, $\sum_{u=0}^{R-1} \overline{p}_{uv} = \frac{1}{S}$, where $\overline{p}_{uv} = \mathbb{P}(U = \frac{1}{R+1} + \frac{u}{R+1}, V = \frac{1}{S+1} + \frac{u}{S+1})$.*

Remark 6.2. A very similar definition is given in [30], who investigated such 'discrete copulas' in the case $R = S$. The 'copula pmf' here coincides essentially with the bistochastic matrix of their Proposition 2. See also [33, 34, 39] and [7, Section 3.1.1]. □

6.3 Existence and uniqueness of the copula pmf

Defining the copula pmf of $\mathbf{p}$ as the member of $[\mathbf{p}]$ with uniform margins raises the question of the existence and uniqueness of such an element on $[\mathbf{p}]$. This question is linked to the algebraic problem of 'matrix scaling': '*Given a nonnegative matrix A, can we find diagonal matrices D_1 and D_2 such that $D_1 A D_2$ is doubly stochastic?*' [48] showed that the answer is affirmative if A is a positive square matrix, a result later generalised to nonnegative and/or non-square matrices; see [25] for a review. This allows a simple necessary and sufficient criterion for the existence and uniqueness of the copula pmf of a given $\mathbf{p} \in \mathcal{P}_{R \times S}$ to be formulated.

When all p_{xy}'s are positive in $\mathbf{p}$, it can directly be deduced from [48, 49] that the copula pmf exists and is unique. Hence the non-trivial case is again when structural zeros are present in $\mathbf{p}$, with their layout being the essential feature. Define the support $\mathrm{Supp}(\mathbf{p}) = \{(x, y) \in \mathcal{S}_X \times \mathcal{S}_Y \text{ s.t. } p_{xy} > 0\}$ and $N(\mathbf{p}) = \left\{ (v_X \times v_Y) : v_X \subset \mathcal{S}_X, v_Y \subset \mathcal{S}_Y \text{ s.t. } \sum_{(x,y) \in v_X \times v_Y} p_{xy} = 0 \right\}$, the set of *rectangular* subsets of $\mathcal{S}_X \times \mathcal{S}_Y$ on which $\mathbf{p}$ is null. See that $N(\mathbf{p}) = \varnothing$ if and only if $\mathrm{Supp}(\mathbf{p}) = \mathcal{S}_X \times \mathcal{S}_Y$, i.e. $p_{xy} > 0 \, \forall (x, y)$.

Let $\mathcal{C}_{R \times S} = \{\mathbf{p} \in \mathcal{P}_{R \times S} : p_{x\bullet} = \frac{1}{R} \, \forall x \in \mathcal{S}_X, p_{\bullet y} = \frac{1}{S} \, \forall y \in \mathcal{S}_Y\} \subset \mathcal{P}_{R \times S}$, the set of all $(R \times S)$-discrete copulas as per Definition 6.1. For any discrete set A, denote $|A|$ the number of elements in A. Then:

Theorem 6.1. *Let $\mathbf{p} \in \mathcal{P}_{R \times S}$.*

(a) *Suppose that, for all $(v_X \times v_Y) \in N(\mathbf{p})$, $\frac{|v_X|}{R} + \frac{|v_Y|}{S} < 1$. Then, there exists a unique $\overline{\mathbf{p}} \in [\mathbf{p}] \cap \mathcal{C}_{R \times S}$;*

(b) *Suppose that, for all $(v_X \times v_Y) \in N(\mathbf{p})$, $\frac{|v_X|}{R} + \frac{|v_Y|}{S} \leq 1$, with $\frac{|\tilde{v}_X|}{R} + \frac{|\tilde{v}_Y|}{S} = 1$ for some $(\tilde{v}_X \times \tilde{v}_Y) \in N(\mathbf{p})$.*

  (i) *If, for all $(\tilde{v}_X \times \tilde{v}_Y) \in N(\mathbf{p})$ such that $\frac{|\tilde{v}_X|}{R} + \frac{|\tilde{v}_Y|}{S} = 1$, $(\mathcal{S}_X \backslash \tilde{v}_X \times \mathcal{S}_Y \backslash \tilde{v}_Y) \in N(\mathbf{p})$, then there exists a unique $\overline{\mathbf{p}} \in [\mathbf{p}] \cap \mathcal{C}_{R \times S}$;*

  (ii) *If there exists $(\tilde{v}_X^\star \times \tilde{v}_Y^\star) \in N(\mathbf{p})$ such that $\frac{|\tilde{v}_X^\star|}{R} + \frac{|\tilde{v}_Y^\star|}{S} = 1$ and $(\mathcal{S}_X \backslash \tilde{v}_X^\star \times \mathcal{S}_Y \backslash \tilde{v}_Y^\star) \notin N(\mathbf{p})$, then $[\mathbf{p}] \cap \mathcal{C}_{R \times S} = \varnothing$ but there exists a unique $\overline{\mathbf{p}} \in \mathrm{Cl}([\mathbf{p}]) \cap \mathcal{C}_{R \times S}$;*

(c) *Suppose that there exists $(\widetilde{\widetilde{v}}_X \times \widetilde{\widetilde{v}}_Y) \in N(\mathbf{p})$ such that $\frac{|\widetilde{\widetilde{v}}_X|}{R} + \frac{|\widetilde{\widetilde{v}}_Y|}{S} > 1$. Then,* $\mathrm{Cl}\left([\mathbf{p}]\right) \cap \mathcal{C}_{R \times S} = \varnothing$.

Proof. Combine Theorems 3.1, 3.2 and 3.3 in [5]. □

Case (a) is the 'easy' case, which covers all $(R \times S)$-distributions with no structural zeros ($N(\mathbf{p}) = \varnothing$), but not only: $\mathbf{p}$ is allowed to have structural zeros ($N(\mathbf{p}) \neq \varnothing$), provided those are not too 'prominent' in the specified sense. The unique $\overline{\mathbf{p}}$ is obviously the copula pmf of $\mathbf{p}$. From (6.4), $\mathrm{Supp}(\overline{\mathbf{p}}) = \mathrm{Supp}(\mathbf{p})$, that is, the pattern of structural zeros (if any) is the same in $\mathbf{p}$ and $\overline{\mathbf{p}}$.

Case (b) is the critical case. In case (b (i)), the matrix $\mathbf{p}$ can be made block-diagonal by some permutations of its rows and columns. Then, each sub-block of non-zero elements of $\mathbf{p}$ can be dealt with separately when adjusting the margins, and it remains possible to write $\overline{\mathbf{p}}$ under the form (6.3), that is $\overline{\mathbf{p}} \in [\mathbf{p}]$, and $\mathrm{Supp}(\overline{\mathbf{p}}) = \mathrm{Supp}(\mathbf{p})$. In the Bernoulli case, this corresponds to 'absolute association' ((i) in (5.7)).

By contrast, in case (b (ii)), the matrix $\mathbf{p}$ cannot be made block-diagonal. For complying with the uniform margins constraint, new zeros must be created in $\overline{\mathbf{p}}$ which must therefore be a limit point of $[\mathbf{p}]$ in the sense (6.5). Then $\overline{\mathbf{p}} \in \mathrm{Cl}([\mathbf{p}])$ and $\mathrm{Supp}(\overline{\mathbf{p}}) \subset \mathrm{Supp}(\mathbf{p})$. The new zeros are created on $\bigcup_{(\tilde{v}_X^\star \times \tilde{v}_Y^\star) \in N(\mathbf{p})} (\mathcal{S}_X \setminus \tilde{v}_X^\star \times \mathcal{S}_Y \setminus \tilde{v}_Y^\star)$. But it holds true that $\Omega(\overline{\mathbf{p}}) = \Omega(\mathbf{p})$ (in the sense of Remark 6.1) and $\overline{\mathbf{p}} \in \mathcal{C}_{R \times S}$, hence $\overline{\mathbf{p}}$ is again the unique copula pmf of $\mathbf{p}$. In the Bernoulli case, this corresponds to 'complete association' ((ii) in (5.7)).

Finally, case (c) establishes the non-existence of a copula pmf when the structural zeros form a bulky subset of $\mathbf{p}$. As the zeros cannot be turned into positive values by (6.3) and are frozen, there do not remain sufficiently many degrees of freedom for adjusting the marginals. Pragmatically, the dependence between X and Y is so overly dictated by the structural zeros that an approach based on odds ratios is pointless.

The above observations allow us to state:

Corollary 6.1 (**Existence and uniqueness of the copula pmf**). *The bivariate discrete distribution $\mathbf{p} \in \mathcal{P}_{R \times S}$ admits a unique copula pmf $\overline{\mathbf{p}}$ if and only if $\frac{|v_X|}{R} + \frac{|v_Y|}{S} \leq 1$ for all $(v_X \times v_Y) \in N(\mathbf{p})$. By definition, the copula pmf $\overline{\mathbf{p}}$ has discrete uniform margins, is such that $\Omega(\overline{\mathbf{p}}) = \Omega(\mathbf{p})$ (in the sense of Remark 6.1) and $\mathrm{Supp}(\overline{\mathbf{p}}) \subseteq \mathrm{Supp}(\mathbf{p})$.*

Finally, the following result provides an interesting characterisation of that copula pmf.

Theorem 6.2. *If the bivariate discrete distribution $\mathbf{p} \in \mathcal{P}_{R \times S}$ admits a copula pmf $\overline{\mathbf{p}}$, then $\overline{\mathbf{p}}$ is the unique element minimising the Kullback-Leibler divergence*

$$D_{KL}(\mathbf{p}^\star \| \mathbf{p}) = \sum_{(x,y) \in \mathrm{Supp}(\mathbf{p})} p_{xy}^\star \log \frac{p_{xy}^\star}{p_{xy}}$$

over the set $\{\mathbf{p}^\star \in \mathcal{C}_{R \times S} : \mathrm{Supp}(\mathbf{p}^\star) \subseteq \mathrm{Supp}(\mathbf{p})\}$.

Proof. See [5], Theorems 3.2 and 3.3. □

The copula pmf $\overline{\mathbf{p}}$ is thus the bivariate discrete distribution with uniform marginals which is the closest to the initial $\mathbf{p}$ in terms of the Kullback-Leibler divergence. Not only this provides a quantitative description of the copula pmf $\overline{\mathbf{p}}$ (as opposed to the rather qualitative definition based on the concept of 'nucleus' in Section 6.2), it also allows a clear parallel to be drawn between the proposed discrete copula and its continuous counterpart. Indeed it is known that a similar characterisation of the copula of a continuous vector exists, under mild conditions, in terms of I-projection [42, 44].

6.4 Iterated proportional fitting procedure

Unlike in the Bernoulli case (5.5), $\overline{\mathbf{p}}$ is usually not available in closed form in the general $(R \times S)$-case. (Specific models lead to closed form copula pmf, though; see Section 7.) However, one can easily extract $\overline{\mathbf{p}}$ from any $\mathbf{p} \in \mathcal{P}_{R \times S}$ by *iterated proportional fitting* (IPF). This consists of alternately normalising the rows and columns of $\mathbf{p}$ to have uniform marginals. In effect, it reproduces (6.5), alternately left- and right-multiplying the current version of $\mathbf{p}$ by diagonal matrices Φ_k and Ψ_k, which leaves all odds-ratios unaffected. Hence IPF perfectly fits in our framework: the output of any iteration remains in $[\mathbf{p}]$. The convergence of IPF was investigated in [14, 26] and [43]; see also [4, Section 3.6] and [41, Section 12.2]. The IPF seeded on any $\mathbf{p} \in \mathcal{P}_{R \times S}$ indeed converges to its copula pmf $\overline{\mathbf{p}}$ provided that it exists, i.e. under the condition of Corollary 6.1. The convergence of the IPF procedure is geometric in case (a) and $(b\,(i))$, and arithmetic in case $(b\,(ii))$ [5]. The R package `mipfp` [3] implements the IPF.

6.5 Construction of arbitrary bivariate discrete distributions with given copula pmf

Similarly to Section 5.6, one may want to construct a $(R \times S)$-discrete distribution with particular marginal distributions $\mathbf{p}_X$ and $\mathbf{p}_Y$ and dependence structure driven by a copula pmf $\overline{\mathbf{p}}$. The existence and uniqueness of such a distribution is (partially) established by the following result, analogue to Theorem 6.1.

Theorem 6.3. *Let $\mathbf{p}_X = (p_{0\bullet}, p_{1\bullet}, \ldots, p_{R-1\bullet})$ and $\mathbf{p}_Y = (p_{\bullet 0}, p_{\bullet 1}, \ldots, p_{\bullet S-1})$ be some target marginal distributions for X and Y, respectively. Let $\overline{\mathbf{p}} \in \mathcal{C}_{R \times S}$, a $(R \times S)$-copula pmf.*

(a) *Suppose that, for all $(\nu_X \times \nu_Y) \in N(\overline{\mathbf{p}})$, $\sum_{x \in \nu_X} p_{x\bullet} + \sum_{y \in \nu_Y} p_{\bullet y} < 1$. Then, there exists a unique $\mathbf{p} \in [\overline{\mathbf{p}}]$ with the requested marginal distributions;*
(b) *Suppose that, for all $(\nu_X \times \nu_Y) \in N(\overline{\mathbf{p}})$, $\sum_{x \in \nu_X} p_{x\bullet} + \sum_{y \in \nu_Y} p_{\bullet y} \leq 1$, with $\sum_{x \in \tilde{\nu}_X} p_{x\bullet} + \sum_{y \in \tilde{\nu}_Y} p_{\bullet y} = 1$ for some $(\tilde{\nu}_X \times \tilde{\nu}_Y) \in N(\overline{\mathbf{p}})$. If, for all $(\tilde{\nu}_X \times \tilde{\nu}_Y) \in N(\overline{\mathbf{p}})$ such that $\sum_{x \in \tilde{\nu}_X} p_{x\bullet} + \sum_{y \in \tilde{\nu}_Y} p_{\bullet y} = 1$, $(\mathcal{S}_X \backslash \tilde{\nu}_X \times \mathcal{S}_Y \backslash \tilde{\nu}_Y) \in N(\overline{\mathbf{p}})$, then there exists a unique $\mathbf{p} \in [\overline{\mathbf{p}}]$ with the requested marginal distributions.*

Proof. See [5], Theorems 3.1 and 3.2. □

This result guarantees the existence of the requested distribution $\mathbf{p}$ in 'easy' cases: no zeros in $\overline{\mathbf{p}}$, or zeros not lying on rows and columns carrying large target marginal weights, or block-diagonal copula pmf $\overline{\mathbf{p}}$. It does not say that such a $\mathbf{p}$ does not exist in the other cases. In fact, such distribution may exist, as evidenced by (5.10) in the Bernoulli case. However, reconstructing $\mathbf{p}$ then is not achieved through the transformation (6.3), as some zeros of $\overline{\mathbf{p}}$ must be turned back into a positive probability, hence $\mathbf{p} \not\sim \overline{\mathbf{p}}$ as $\mathbf{p}$ belongs to an orbit of which $\overline{\mathbf{p}}$ is only a limit point. Existence and uniqueness of $\mathbf{p}$ in those cases remains an open question; however, a geometric perspective similar to Remark 5.2 suggests positive conclusions.

6.6 Yule's coefficient

By analogy with Section 5.5, one can define a margin-free measure of overall concordance in $\mathbf{p}$ as Pearson's correlation coefficient computed on $\overline{\mathbf{p}}$. We call such a coefficient Yule's coefficient Υ, which can again be regarded as the discrete analogue to Spearman's ρ. Suppose that U is discrete uniform on $\{\frac{1}{R+1}, \frac{2}{R+1}, \ldots, \frac{R}{R+1}\}$, V is discrete uniform on $\{\frac{1}{S+1}, \frac{2}{S+1}, \ldots, \frac{S}{S+1}\}$, and their joint pmf is given by the copula pmf $\overline{\mathbf{p}}$. It can then be checked that Pearson's correlation between U and V is

$$\Upsilon = 3\sqrt{\frac{(R-1)(S-1)}{(R+1)(S+1)}} \left(\frac{4}{(R-1)(S-1)} \sum_{u=0}^{R-1} \sum_{v=0}^{S-1} uv\bar{p}_{uv} - 1 \right). \tag{6.6}$$

This coefficient is equal to 1 or -1 if and only if $\overline{\mathbf{p}}$ is a diagonal matrix, which obviously requires $R = S$. The diagonal copula pmf's of size $(R \times R)$ are clearly

$$\overline{\mathbf{m}} \doteq \begin{pmatrix} \frac{1}{R} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{R} \end{pmatrix} \quad \text{or} \quad \overline{\mathbf{w}} \doteq \begin{pmatrix} 0 & & \frac{1}{R} \\ & \cdot^{\cdot^{\cdot}} & \\ \frac{1}{R} & & 0 \end{pmatrix}, \tag{6.7}$$

the Fréchet bounds analogous to (5.6). Note that any $\mathbf{p} \in \mathcal{P}_{R \times R}$ represented by a diagonal matrix is easily seen to admit $\overline{\mathbf{m}}$ or $\overline{\mathbf{w}}$ as copula pmf. Those fall into case $(b\,(i))$ of Theorem 6.1 and correspond to (positive or negative) 'absolute association'; cf. Section 5.4. There also exist non-diagonal distributions $\mathbf{p} \in \mathcal{P}_{R \times R}$, belonging to case $(b\,(ii))$ of Theorem 6.1, which admit $\overline{\mathbf{m}}$ or $\overline{\mathbf{w}}$ as copula pmf as well. Those would be akin to 'complete association'. As in the Bernoulli case, the distinction between 'absolute' and 'complete association' is only a marginal feature which must be ignored by the copula. For instance, positive 'absolute association', i.e. probability weight concentrated on the main diagonal of $\mathbf{p}$, is only possible if $\mathbf{p}_X \equiv \mathbf{p}_Y$. By contrast, dependence as strong as can be between unequal discrete marginals must turn into 'complete association'. As a result, $\Upsilon = \pm 1$ without distinction between 'absolute' and 'complete' association.

Now, when $R \neq S$, $|\Upsilon|$ cannot reach 1. Indeed, if X and Y do not take the same number of values, the associated copula pmf can never approach any of the diagonal forms $\overline{\mathbf{m}}$ or $\overline{\mathbf{w}}$ for the obvious reason, preventing any sense of 'perfect dependence'. In that case, the maximum value attained by $|\Upsilon|$ occurs when $\mathbf{p}$ is the pmf associated to the Fréchet bounds in the class of $(R \times S)$-bivariate discrete distributions with uniform margins [16, 'Exemple I']. It is also clear that, like Spearman's ρ, Υ only detects monotonic dependence ('concordance') between X and Y. In particular, for $\max(R, S) > 2$, Υ can be 0 even when X and Y are not independent. Genuine measures of dependence Δ, in the sense of $\Delta = 0 \iff X \perp Y$, may be defined along the same way as in [17].

Example 6.1. The copula pmf being, by definition, margin-free, its construction only requires a sense of order for the 'values' of X and Y. In particular, if X and/or Y are ordinal random variables, then it remains meaningful to construct their copula pmf in order to understand their dependence. This is illustrated here through data on congenital sex organ malformations cross-classified by maternal alcohol consumption from a study described in [22]:

	Maternal Alcohol Consumption (drinks/day)					
	0	< 1	1 – 2	3 – 5	≥ 6	
No Malformation	17, 066	14, 464	788	126	37	32, 481 ·
Malformation	48	38	5	1	1	93
	17, 114	14, 502	793	127	38	32, 574

Agreed that 'No malformation' $\prec$ 'Malformation', the dependence between maternal alcohol consumption and congenital malformation can be understood through that of the bivariate discrete distribution

$x \backslash Y$	0	1	2	3	4	
0	0.52391	0.44404	0.02419	0.00387	0.00114	0.99714
1	0.00147	0.00117	0.00015	0.00003	0.00003	0.00286
	0.52539	0.44520	0.02434	0.00390	0.00117	1

$$\tag{6.8}$$

The IPF procedure (Section 6.4) returns the copula pmf of this distribution:

$$\begin{pmatrix} 0.137 & 0.140 & 0.098 & 0.087 & 0.037 \\ 0.063 & 0.060 & 0.102 & 0.113 & 0.163 \end{pmatrix},$$

shown as a confetti plot in Figure 6.1. The adverse effect of maternal alcohol consumption on the risk of congenital malformation appears clearly, and is quantified by a positive value of Yule's coefficient of $\Upsilon = 0.358$. Entirely margin-free, such a copula-based measure of association between two ordinal random variables
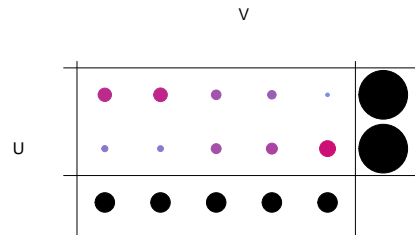
Figure 6.1: Confetti plot of the copula pmf $\overline{\mathbf{p}}$ for distribution (6.8).

does not rely on assigning scores to each category as is otherwise necessary [28, Section 2.3] – the 'labels' $X \in \{0, 1\}$ and $Y \in \{0, 1, 2, 3, 4\}$ in (6.8) have no impact whatsoever. This seems desirable, as [21, p. 740] noted: "*We feel that the use of arbitrary scores to motivate measures is infrequently appropriate*."

7 Parametric discrete copulas

Paralleling the continuous case, one can construct parametric models of copula pmf's. In fact, any parametric continuous copula readily gives rise to a discrete copula pmf of any dimension ($R \times S$), as described in Section 7.1. One may also think of specific discrete copulas originating from particular bivariate discrete distributions, such as the Binomial copula (Section 7.2) or truncated Geometric copula (Section 7.3).

7.1 Discrete versions of classical continuous copulas

Let C be a continuous copula as in Definition 1.1. For $u \in \{0, \ldots, R - 1\}$ and $v \in \{0, \ldots, S - 1\}$, define

$$\overline{p}_{uv} = C\left(\frac{u+1}{R}, \frac{v+1}{S}\right) - C\left(\frac{u}{R}, \frac{v+1}{S}\right) - C\left(\frac{u+1}{R}, \frac{v}{S}\right) + C\left(\frac{u}{R}, \frac{v}{S}\right). \tag{7.1}$$

Then, as C has uniform margins on $\mathbb{J}$, it follows, for any u, v, $\sum_{v=0}^{S-1} \overline{p}_{uv} = \frac{1}{R}$ and $\sum_{u=0}^{R-1} \overline{p}_{uv} = \frac{1}{S}$. Hence the $(R \times S)$-discrete distribution $\overline{\mathbf{p}} = [\overline{p}_{uv}]_{\substack{u=0,\ldots,R-1 \\ v=0,\ldots,S-1}}$ is a copula pmf as defined by Definition 6.1.

For simple parametric continuous copulas C, such $\overline{\mathbf{p}}$ can be written in closed form. For instance, it can be checked that the $(R \times S)$-discrete version of the FGM copula is, for $\theta \in [-1, 1]$,

$$\overline{p}_{uv} = \frac{1}{RS}\left(1 + \theta(1 - \frac{2u+1}{R})(1 - \frac{2v+1}{S})\right), \qquad (u, v) \in \{0, \ldots, R - 1\} \times \{0, \ldots, S - 1\}. \tag{7.2}$$

For $\theta = 0$ in (7.2), one finds $\overline{p}_{uv} = 1/RS$ for all (u, v), which is the $(R \times S)$-independence copula pmf:

$$\overline{\mathbf{p}} \doteq \overline{\boldsymbol{\pi}} = \frac{1}{RS}\mathbf{1}_R\mathbf{1}_S^T, \qquad \text{with } \mathbf{1}_k = \underbrace{(1, \ldots, 1)}_{k \text{ times}}^T. \tag{7.3}$$

Remark 7.1. The discrete copula pmf's derived from a continuous one through (7.1) are obtained by overlaying C on the regular mesh $\{0, \frac{1}{R}, \ldots, \frac{R-1}{R}, 1\} \times \{0, \frac{1}{S}, \ldots, \frac{S-1}{S}, 1\}$ over the unit square $\mathbb{J}$. The discretisation is thus carried out 'in the copula world', keeping all marginals uniform. The so-produced discrete copula pmf's $\overline{\mathbf{p}}$ can then be used in a second time for modelling dependence and/or constructing new bivariate discrete distributions with prescribed marginals. The idea of two distinct building blocks, the marginals on one side and the dependence/copula on the other, is maintained. By contrast, when writing (1.1) for a bivariate discrete distribution F_{XY} with a certain continuous copula C, the discretisation is achieved by overlaying C on the mesh $\text{Ran}\,F_X \times \text{Ran}\,F_Y$ set by the margins of F_{XY}. So the discretised copula is defined by the margins,

which explains why dependence and marginal distributions can never be separated in such models. The two approaches coincide for a continuous vector (X, Y), though, as then the mesh Ran $F_X \times$ Ran F_Y reduces down to the whole unit square $\mathbb{J}$, akin to a 'continuous regular mesh'.

7.2 The Binomial copula

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of a bivariate Bernoulli random variable with pmf (5.1). [31, Section 3] defined the bivariate Binomial as the distribution of the vector $(X, Y) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right)$. Its pmf is, for $(x, y) \in \{0, \ldots, n\} \times \{0, \ldots, n\}$,

$$\mathbb{P}(X = x, Y = y) = \sum_{k=\max(x+y-n,0)}^{\min(x,y)} \binom{n}{k, x-k, y-k, n-x-y+k} p_{00}^{n-x-y+k} p_{10}^{x-k} p_{01}^{y-k} p_{11}^k.$$

Then, it can be checked that the odds-ratios (6.1) are

$$\omega_{xy} = \sum_{k=\max(x+y-n,0)}^{\min(x,y)} \frac{\binom{n}{k, x-k, y-k, n-x-y+k}}{\binom{n}{x}\binom{n}{y}} \omega^k,$$

where ω is the odds-ratio of the initial bivariate Bernoulli (3.1). For n fixed, the dependence structure in a bivariate Binomial is thus only driven by one parameter ω, and the corresponding Binomial(n)-copula, which is a $((n+1) \times (n+1))$-discrete distribution with uniform margins, is a one-parameter model.

For instance, if $n = 2$, the bivariate Binomial distribution and its odds ratio matrix (6.2) are, respectively,

$$\mathbf{p} = \begin{pmatrix} p_{00}^2 & 2p_{00}p_{01} & p_{01}^2 \\ 2p_{00}p_{10} & 2(p_{11}p_{00} + p_{10}p_{01}) & 2p_{11}p_{01} \\ p_{10}^2 & 2p_{10}p_{11} & p_{11}^2 \end{pmatrix} \in \mathcal{P}_{3\times3}, \quad \text{and} \quad \Omega(\mathbf{p}) = \begin{pmatrix} \frac{1}{2}(\omega+1) & \omega \\ \omega & \omega^2 \end{pmatrix}.$$

Through some algebra one can make the margins of $\mathbf{p}$ into uniforms through (6.3), and one obtains

$$\overline{\mathbf{p}} = \frac{1}{3} \begin{pmatrix} \frac{\omega(\omega+1)}{\omega^2+\omega+1+\sqrt{\omega(\omega+2)(2\omega+1)}} & \frac{\sqrt{\omega(\omega+2)(2\omega+1)}-3\omega}{(\omega-1)^2} & \frac{\omega+1}{\omega^2+\omega+1+\sqrt{\omega(\omega+2)(2\omega+1)}} \\ \frac{\sqrt{\omega(\omega+2)(2\omega+1)}-3\omega}{(\omega-1)^2} & \frac{\omega^2+4\omega+1-2\sqrt{\omega(\omega+2)(2\omega+1)}}{(\omega-1)^2} & \frac{\sqrt{\omega(\omega+2)(2\omega+1)}-3\omega}{(\omega-1)^2} \\ \frac{\omega+1}{\omega^2+\omega+1+\sqrt{\omega(\omega+2)(2\omega+1)}} & \frac{\sqrt{\omega(\omega+2)(2\omega+1)}-3\omega}{(\omega-1)^2} & \frac{\omega(\omega+1)}{\omega^2+\omega+1+\sqrt{\omega(\omega+2)(2\omega+1)}} \end{pmatrix}$$

for $\omega \neq 1$. For $\omega = 1$, of course, $\overline{\mathbf{p}} = \overline{\boldsymbol{\pi}}$, the (3×3)-independence copula pmf (7.3). See also that, for $\omega = 0$ or $\omega = \infty$, $\overline{\mathbf{p}} = \overline{\mathbf{w}}$ and $\overline{\mathbf{p}} = \overline{\mathbf{m}}$, the Fréchet lower and upper bounds (6.7) in 3 dimensions. One also has

$$\Upsilon = \frac{\omega^2 - 1}{\omega^2 + \omega + 1 + \sqrt{\omega(\omega + 2)(2\omega + 1)}}$$

as Yule's coefficient (6.6) for this copula pmf, which is $\Upsilon = -1$ for $\omega = 0$ and $\Upsilon = 1$ for $\omega = \infty$. This family of Binomial copulas is thus comprehensive as it allows all values for Yule's coefficients from -1 and 1.

7.3 The truncated geometric copula

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ again be a sequence of independent replications from the bivariate Bernoulli distribution (5.1). [31, Section 6] defined the bivariate Geometric distribution as the distribution of the vector (X, Y) where X is the number of 0's before the first 1 in the sequence $X_1, X_2, \ldots$, and Y the number of 0's before the first 1 in the sequence $Y_1, Y_2, \ldots$. The pmf is

$$\mathbb{P}(X = x, Y = y) = \begin{cases} p_{00}^x p_{10} p_{\bullet 0}^{y-x-1} p_{\bullet 1} & 0 \leq x < y \\ p_{00}^x p_{11} & 0 \leq x = y \\ p_{00}^y p_{01} p_{0\bullet}^{x-y-1} p_{1\bullet} & 0 \leq y < x \end{cases}. \tag{7.4}$$

Now, consider the truncated random variables $\tilde{X}_N = \min(X, N-1)$ and $\tilde{Y}_N = \min(Y, N-1)$, for some $N \geq 2$. The random vector $(\tilde{X}_N, \tilde{Y}_N)$ has pmf

$$\mathbb{P}(\tilde{X}_N = x, \tilde{Y}_N = y) = \begin{cases} \mathbb{P}(X = x, Y = y) & 0 \leq x, y < N-1 \\ p_{00}^y p_{01} p_{0\bullet}^{N-y-2} & x = N-1, 0 \leq y < N-1 \\ p_{00}^x p_{10} p_{\bullet 0}^{N-x-2} & 0 \leq x < N-1, y = N-1 \\ p_{00}^{N-1} & x = N-1, y = N-1 \end{cases}. \tag{7.5}$$

For instance, for $N = 3$, one obtains a distribution $\mathbf{p}$ equal to

$\tilde{X}_N \backslash \tilde{Y}_N$	0	1	2	
0	p_{11}	$p_{10}p_{\bullet 1}$	$p_{10}p_{\bullet 0}$	$p_{1\bullet}$
1	$p_{01}p_{1\bullet}$	$p_{00}p_{11}$	$p_{00}p_{10}$	$p_{0\bullet}p_{1\bullet}$
2	$p_{01}p_{0\bullet}$	$p_{00}p_{01}$	p_{00}^2	$p_{0\bullet}^2$
	$p_{\bullet 1}$	$p_{\bullet 0}p_{\bullet 1}$	$p_{\bullet 0}^2$	

whose odds ratio matrix (6.2) is

$$\Omega(\mathbf{p}) = \begin{pmatrix} \omega \frac{p_{11}}{p_{1\bullet}\cdot p_{\bullet 1}} & \omega \frac{p_{10}}{p_{1\bullet}\cdot p_{\bullet 0}} \\ \omega \frac{p_{01}}{p_{0\bullet}\cdot p_{\bullet 1}} & \omega \frac{p_{00}}{p_{0\bullet}\cdot p_{\bullet 0}} \end{pmatrix}, \tag{7.6}$$

where ω is again the odds-ratio of the initial bivariate Bernoulli. As opposed to the binomial case, the structure of dependence in this bivariate Geometric depends not only on ω, but also on the marginals of the initial bivariate Bernoulli through $p_{\bullet 1}$ and $p_{1\bullet}$ (the other quantities in (7.6) are functions of ω, $p_{\bullet 1}$ and $p_{1\bullet}$; Section 5.6). It is thus a 3-parameters copula family.

One can define a 'standard' version of it by fixing $p_{\bullet 1} = p_{1\bullet} = 1/2$, that is, assuming that the initial bivariate Bernoulli distribution is the copula pmf (5.5). Then one recovers a one parameter copula model, for which one obtains (after some algebra) the explicit form of the copula pmf:

$$\overline{\mathbf{p}} = \frac{1}{3} \begin{pmatrix} \frac{2\omega}{2\omega + \sqrt{8\omega+1}+1} & \frac{\sqrt{8\omega+1}+1}{2(2\omega+\sqrt{8\omega+1}+1)} & \frac{\sqrt{8\omega+1}+1}{2(2\omega+\sqrt{8\omega+1}+1)} \\ \frac{\sqrt{8\omega+1}+1}{2(2\omega+\sqrt{8\omega+1}+1)} & \frac{\sqrt{\omega}(\sqrt{8\omega+1}+1)^2}{4(\sqrt{\omega}+1)(2\omega+\sqrt{8\omega+1}+1)} & \frac{4\omega-1-\sqrt{8\omega+1}}{4(\sqrt{\omega}-1)(\sqrt{\omega}+1)^2} \\ \frac{\sqrt{8\omega+1}+1}{2(2\omega+\sqrt{8\omega+1}+1)} & \frac{4\omega-1-\sqrt{8\omega+1}}{4(\sqrt{\omega}-1)(\sqrt{\omega}+1)^2} & \frac{\sqrt{\omega}(\sqrt{8\omega+1}+1)^2}{4(\sqrt{\omega}+1)(2\omega+\sqrt{8\omega+1}+1)} \end{pmatrix} \tag{7.7}$$

for $\omega \neq 1$. If $\omega = 1$, then $\overline{\mathbf{p}} = \overline{\boldsymbol{\pi}}$. If $\omega = \infty$, then $\overline{\mathbf{p}} = \overline{\mathbf{m}}$, and $\Upsilon = 1$. On the other hand, when $\omega = 0$, then

$$\overline{\mathbf{p}} = \begin{pmatrix} 0 & 1/6 & 1/6 \\ 1/6 & 0 & 1/6 \\ 1/6 & 1/6 & 0 \end{pmatrix},$$

with $\Upsilon = -1/2$ in that case. We call (7.7) the truncated Geometric(3) copula pmf. Any threshold value $N \geq 2$ can be considered as well, yielding the truncated Geometric(N) copula pmf for $(N \times N)$-distributions.

8 Discrete distributions with infinite support

8.1 General case

The above construction carries over to the case of discrete random variables with infinite support, say $\mathbb{N}$ (without loss of generality, as any infinite discrete set is in one-to-one correspondence with $\mathbb{N}$). A way of approaching this is to first consider truncated versions of the random variables of interest, like in Section 7.3, and then let the truncation threshold N tend to ∞. Consider a bivariate discrete vector (X, Y) supported on $\mathbb{N} \times \mathbb{N}$, and assume here that $p_{xy} = \mathbb{P}(X = x, Y = y) > 0, \forall (x, y)$. Let $\tilde{X}_N = \min(X, N-1)$ and $\tilde{Y}_N = \min(Y, N-1)$,

for some $N \geq 2$. The pmf of $(\tilde{X}_N, \tilde{Y}_N)$ is

$$\tilde{p}_{N;xy} = \begin{cases} p_{xy} & \text{if } 0 \leq x, y < N - 1 \\ \sum_{y^* \geq N} p_{xy^*} & \text{if } 0 \leq x < N, y = N - 1 \\ \sum_{x^* \geq N} p_{x^*y} & \text{if } 0 \leq y < N, x = N - 1 \\ \sum_{x^* \geq N} \sum_{y^* \geq N} p_{x^*y^*} & \text{if } x = N - 1, y = N - 1 \\ 0 & \text{if } x \geq N, y \geq N \end{cases}.$$

Denote $\tilde{\mathbf{p}}_N$ the corresponding matrix in $\mathcal{P}_{N \times N}$. For all integer N, it follows from Corollary 6.1 that this bivariate discrete distribution admits a unique copula pmf $\bar{\tilde{\mathbf{p}}}_N \in [\tilde{\mathbf{p}}_N] \cap \mathcal{C}_{N \times N}$, as all $\tilde{p}_{N;xy}$ are positive on $\{0, \ldots, N-1\} \times \{0, \ldots, N-1\}$. By Definition 6.1, the copula pmf $\bar{\tilde{\mathbf{p}}}_N$ is the pmf of a vector $(\tilde{U}_N, \tilde{V}_N)$ whose both margins are Discrete uniform on $\{\frac{1}{N+1}, \frac{2}{N+1}, \ldots, \frac{N}{N+1}\}$. Now, let $N \to \infty$. Then, $\tilde{U}_N \xrightarrow{\mathcal{L}} \mathcal{U}_{[0,1]}$ and $\tilde{V}_N \xrightarrow{\mathcal{L}} \mathcal{U}_{[0,1]}$, so that $(\tilde{U}_N, \tilde{V}_N)$ converges in law to a bivariate distribution with continuous uniform marginals, that is, a (continuous) copula as per Definition 1.1; the proof is given in [30, Theorems 1 and 2] and [7, Theorem 3.1.8]. So, the dependence structure of a discrete bivariate vector supported on $\mathbb{N} \times \mathbb{N}$ can be represented by a *unique* continuous copula. Yet, this unique copula is not any of the copulas C satisfying (1.1).

Indeed, analogously to Remark 7.1, (1.1) reconstructs the bivariate discrete distribution F_{XY} by overlaying a copula C on the mesh Ran $F_X \times$ Ran F_Y over $\mathcal{I}$. Such copula is not unique and is indissociable to the margins. By contrast, the above construction singles out one unique copula which represents the 'core' of F_{XY} in the spirit of the marginal transformations described in Section 6.2. It is independent of the margins, as it is a representation of all the odds ratios ω_{xy} (6.1) for $(x, y) \in \mathbb{N}_+ \times \mathbb{N}_+$. The bivariate discrete distribution F_{XY} can thus be broken down into its marginal distributions on one hand, and its unique copula on the other, like in the continuous case. A difference is that here, the combination of the copula and the marginals is not carried out by (1.1), but by a continuous version of IPF [25, Sections 6.3 and 6.4, and references therein].

8.2 The Geometric copula

Consider the truncated Geometric distribution $\tilde{\mathbf{p}}_N$ given by (7.5), and set $p_{1\bullet} = p_{\bullet 1} = 1/2$. For any $N \geq 2$, one gets a copula pmf's $\bar{\tilde{\mathbf{p}}}_N \in \mathcal{C}_{N \times N}$ involving only one parameter ω, like (7.7) for $N = 3$. As $N \to \infty$, those discrete pmf's turn into a continuous distribution with uniform margins as pictured in Figure 8.1 (left: $\omega = 2$; right: $\omega = 1/2$). For $\omega > 1$, this copula admits a singularity along the main diagonal of the unit square $\mathcal{I}$. This is reminiscent of the Marshall-Olkin copula [37, Section 3.1.1], a link to which could have been expected here given that the Marshall-Olkin bivariate Exponential distribution is the limit version of the bivariate Geometric distribution introduced above [31, Section 6]. The Geometric copula, however, remains a representative of the inner dependence structure in the discrete vector (X, Y) whose pmf is (7.4), and is not the Marshall-Olkin copula as appears clearly when $\omega < 1$. Then, the limiting Geometric copula density is seen to be identically null on the main diagonal of $\mathcal{I}$, forming some sort of 'inverse singularity'.
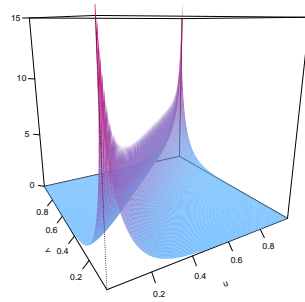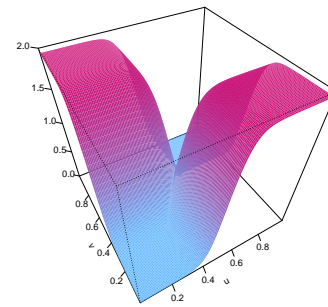
8.3 The Poisson copula

Let $Z_{10} \sim \mathcal{P}(\lambda_{10})$, $Z_{01} \sim \mathcal{P}(\lambda_{01})$ and $Z_{11} \sim \mathcal{P}(\lambda_{11})$ be three independent Poisson random variables, with $\lambda_{10}, \lambda_{01} > 0$ and $\lambda_{11} \geq 0$. Then define

$$X = Z_{10} + Z_{11} \qquad \text{and} \qquad Y = Z_{01} + Z_{11}. \tag{8.1}$$

The distribution of the vector (X, Y) is classically known as the bivariate Poisson distribution, parameterised by the triplet $(\lambda_{10}, \lambda_{01}, \lambda_{11})$. Its pmf is [31, Section 4]:

$$p_{xy} = \mathbb{P}(X = x, Y = y) = e^{-(\lambda_{10} + \lambda_{01} + \lambda_{11})} \frac{\lambda_{10}^x}{x!} \frac{\lambda_{01}^y}{y!} \sum_{i=0}^{\min(x,y)} i! \binom{x}{i} \binom{y}{i} \left(\frac{\lambda_{11}}{\lambda_{10}\lambda_{01}}\right)^i,$$

(a) $\omega = 2$ **(b)** $\omega = 1/2$

Figure 8.1: The Geometric copula density.

for $(x, y) \in \mathbb{N} \times \mathbb{N}$, and $X \sim \mathcal{P}(\lambda_{10} + \lambda_{11})$ and $Y \sim \mathcal{P}(\lambda_{01} + \lambda_{11})$. The odds ratios (6.1) reduce down to

$$\omega_{xy} = \sum_{i=0}^{\min(x,y)} i! \binom{x}{i} \binom{y}{i} \left(\frac{\lambda_{11}}{\lambda_{10}\lambda_{01}} \right)^i, \qquad (x, y) \in \mathbb{N}_+ \times \mathbb{N}_+.$$

It is seen that the dependence structure in such a bivariate Poisson vector only depends on the parameter $\omega \doteq \lambda_{11}/(\lambda_{10}\lambda_{01})$. If the bivariate Poisson distribution is understood as a limiting version of a bivariate Binomial [31, Section 4], then this ω would indeed be akin to the odds ratio in the constituting initial bivariate Bernoulli distribution. Acting as in the previous section, one can first truncate X and Y at $N - 1$, for obtaining discrete copula pmf's and then let N tend to infinity for obtaining the Poisson copula densities shown in Figure 8.2 for $\omega = 0.01$ (left) and $\omega = 0.2$ (right).
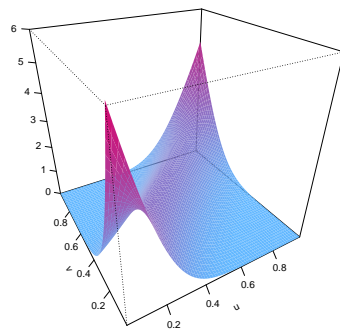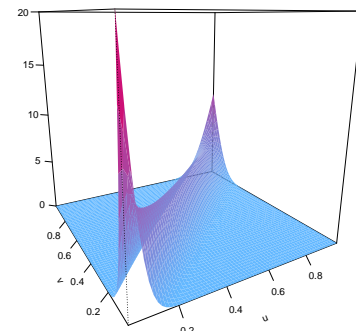
(a) $\omega = 0.01$ **(b)** $\omega = 0.2$

Figure 8.2: The Poisson copula density.

Like in any bivariate discrete distribution built on such an idea of 'trivariate reduction' (8.1), the components X and Y of a bivariate Poisson vector can only show positive association. How to construct bivariate discrete distributions with Poisson marginals showing negative association has been a challenging problem for a long

time. For instance, [23] noted: "*we have been unable to discover explicit in the literature any examples of bivariate Poisson distributions in which the correlation is negative.*" [40] proposed a classical copula construction based on (1.1), thus subject to caution following the discussion in Section 2, in particular, the impossibility of ever disjointing margins and dependence.

By contrast, one can couple any two Poisson distributions with any continuous copula through IPF (Sections 6.4-6.5). Figure 8.3 shows confetti plots of three bivariate discrete distributions with Poisson $\mathcal{P}(2)$ marginals and negative association; coupled through (*a*) a Clayton copula with $\theta = -0.2$; (*b*) a Gaussian copula with $\rho = -0.8$; and (*c*) a Geometric copula with $\omega = 1/2$ (Figure 8.1). This illustrates that the proposed discrete copula approach shares with its continuous counterpart the same flexibility for constructing 'new' bivariate distributions with arbitrary marginals and dependence structure.
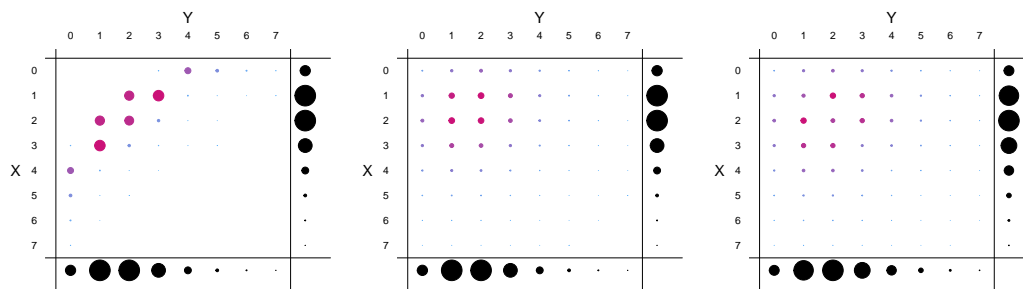

Figure 8.3: Confetti plots of three bivariate discrete distributions with $\mathcal{P}(2)$-margins coupled by (*a*) a Clayton copula with $\theta = -0.2$ (left); (*b*) a Gaussian copula with $\rho = -0.8$ (middle) and (*c*) a Geometric copula with $\omega = 1/2$ (right). All three show negative association between X and Y.

9 Concluding remarks

The classical definition of a copula (Definition 1.1) stems implicitly but manifestly from the Probability Integral Transform result. Hence it is fundamentally grounded in the continuous framework, and there is little surprise that classical copula ideas lead to many inconsistencies when applied on discrete random vectors. What may appear surprising is that a large part of the previous literature in the field has tried to make such an inherently continuous concept forcibly fit the discrete case as well, in spite of those inconsistencies.

In this paper it is argued that the very essence of a copula should not be imprisoned in Definition 1.1. Fundamentally, a copula is akin to an equivalence class of distributions sharing the same dependence structure. Defining such equivalence classes, called *nuclei*, does not require resorting to PIT and hence smoothly carries over to the discrete case. This paper describes that 'discrete copula' construction. All the pleasant properties of copulas for modelling dependence are maintained in the presented discrete framework, such as margin-freeness of anything copula-based or flexibility in constructing bivariate distributions with arbitrary marginals and dependence structure.

Existence and uniqueness of the copula probability mass function, analogue to the copula density in the continuous case, are established under mild conditions. The ideas are first introduced in the bivariate Bernoulli case, and then generalised to distributions supported on $\{0, 1, \ldots, R\} \times \{0, 1, \ldots, S\}$, for some finite R and S, and finally to bivariate distributions supported on $\mathbb{N} \times \mathbb{N}$. Interestingly, the dependence structure in such a ($\mathbb{N} \times \mathbb{N}$)-supported distribution may still be captured by a classical continuous copula, and that copula is unique. However, that copula is *not* to be understood through Sklar's theorem (1.1). The construction gives rise to new continuous copulas, such as the Geometric copula and the Poisson copula, representing the de-

pendence structure in bivariate Geometric and bivariate Poisson distributions, respectively. Purely discrete copulas are also introduced, such as the Binomial copula.

Finally, we note that it is straightforward to generalise the bivariate concepts expounded in this paper to higher dimensions. Definition 6.1 (discrete copula) can be formulated in terms of a multi-dimensional array with uniform margins in every dimension. Higher-order odds-ratios keep their valuable properties, in particular margin-freeness [41, Section 6.2]. The matrix scaling problem is well-studied for higher-dimensional arrays as well [25, Section 6.1, and references therein], and the multi-dimensional IPF algorithm is implemented in the R `mipfp` package [3].
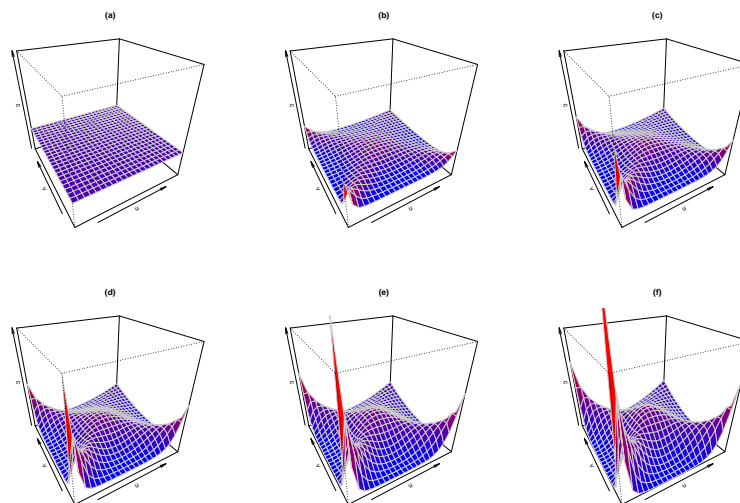
Figure 9.1: Densities of copulas satisfying (A.1), with C_1 the Gaussian copula with $\rho = -0.9$ and C_2 the Clayton copula with parameter $\theta = 3.6$. (a) $\alpha = \beta = 0$ (independence copula); (b) $\alpha = 0.103$, $\beta = 0.097$; (c) $\alpha = 0.206$, $\beta = 0.194$; (b) $\alpha = 0.309$, $\beta = 0.291$; (b) $\alpha = 0.413$, $\beta = 0.387$; (b) $\alpha = 0.516$, $\beta = 0.484$.

A Appendix

Let C_1 and C_2 be two copulas such that $C_1 \prec \Pi \prec C_2$ (concordance ordering; that is, $C_1(u, v) \le uv \le C_2(u, v)$, $\forall (u, v) \in \mathfrak{I}$). Consider the mixture copula defined as

$$C_{\alpha,\beta}(u, v) = \alpha C_1(u, v) + (1 - \alpha - \beta)\Pi(u, v) + \beta C_2(u, v),$$

for $\alpha, \beta \in [0, 1]$ with $\alpha + \beta \le 1$. Call

$$\xi \doteq \frac{\Pi(1 - \pi_X, 1 - \pi_Y) - C_1(1 - \pi_X, 1 - \pi_Y)}{C_2(1 - \pi_X, 1 - \pi_Y) - C_1(1 - \pi_X, 1 - \pi_Y)}$$

and assume $\xi \in (0, 1)$. Then it can be checked that, for any $\alpha \le \xi$ and $\beta = \frac{1-\xi}{\xi} \alpha$, the copula $C_{\alpha,\beta}$ is such that

$$C_{\alpha,\beta}(1 - \pi_X, 1 - \pi_Y) = (1 - \pi_X)(1 - \pi_Y), \tag{A.1}$$

and thus satisfies Sklar's theorem for two independent Bernoulli random variables $X \sim \text{Bern}(\pi_X)$ and $Y \sim \text{Bern}(\pi_Y)$. For any C_1, C_2, the independence copula Π corresponds to $(\alpha, \beta) = (0, 0)$, but there exist infinitely

many other copulas, of various natures and shapes, which satisfy (A.1). For example, for $\pi_X = \pi_Y = 1/2$, the 6 copula densities shown in Figure 9.1 are equally valid for representing the independence between X and Y. Yet, their appearances – and consequently any qualitative or quantitative assessment of the underlying dependence based on them – are dramatically different.

References

[1]   Agresti, A. (2013). *Categorical Data Analysis*. Third Edition. Wiley, Hoboken NJ.
[2]   Altham, P.M. (1970). The measurement of association of rows and columns for an $r \times s$ contingency table. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 32*(1), 63–73.
[3]   Barthélemy, J. and T. Suesse (2018). mipfp: An R package for multidimensional array fitting and simulating multivariate Bernoulli distributions. *J. Stat. Softw. 86*, 2.
[4]   Bishop, Y.M.M., Fienberg, S.E., and P.W. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge.
[5]   Brossard, J. and C. Leuridan. (2018). Iterated proportional fitting procedure and infinite products of stochastic matrices. In Donati-Martin C., Lejay A. and A. Rouault (Eds.), *Séminaire de Probabilités XLIX*, pp. 75–117. Springer, Cham.
[6]   Denuit, M. and P. Lambert (2005). Constraints on concordance measures in bivariate discrete data. *J. Multivariate Anal. 93*(1), 40–57.
[7]   Durante, F. and C. Sempi (2015). *Principles of Copula Theory*. CRC Press, Boca Raton FL.
[8]   Edwards, A.W.F. (1963). The measure of association in a 2×2 table. *J. Roy. Statist. Soc. Ser. A Gen. 126*(1), 109–114.
[9]   Eie, M. and S.-T. Chang (2010). *A Course on Abstract Algebra*. World Scientific, Singapore.
[10] Embrechts, P. (2009). Copulas: a personal view. *J. Risk Ins. 76*(3), 639–650.
[11] Faugeras, O.P. (2015). Maximum coupling of empirical copulas for discrete vectors. *J. Multivariate Anal. 137*, 179–186.
[12] Faugeras, O.P. (2017). Inference for copula modeling of discrete data: a cautionary tale and some facts. *Depend. Model. 5*, 121–132.
[13] Fienberg, S.E. (1968). The geometry of $r \times c$ contingency table. *Ann. Math. Stat. 39*(4), 1186–1190.
[14] Fienberg, S.E. (1970). An iterative procedure for estimation in contingency tables. *Ann. Math. Stat. 41*(3), 907–917.
[15] Fienberg, S.E. and J.P. Gilbert (1970). The geometry of a two by two contingency table. *J. Amer. Statist. Assoc. 65*(330), 694–701.
[16] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon A, Sér. 3*(14), 53-77.
[17] Geenens, G. and P. Lafaye de Micheaux (2020). The Hellinger correlation. *J. Amer. Statist. Assoc.*, in press. Available at https://doi.org/10.1080/01621459.2020.1791132.
[18] Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *Astin Bull. 37*(2), 475–515.
[19] Genest, C., Nešlehová, J., and B. Rémillard (2014). On the empirical multilinear copula process for count data. *Bernoulli 20*(3), 1344–1371.
[20] Genest, C., Nešlehová, J., and B. Rémillard (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivariate Anal. 159*, 82–110.
[21] Goodman, L.A. and W.H. Kruskal (1954). Measures of association for cross classifications. *J. Amer. Statist. Assoc. 49*(268), 732–764.
[22] Graubard, B.I., and E.L. Korn (1987). Choice of column scores for testing independence in ordered $2 \times K$ contingency tables. *Biometrics 43*(2), 471–476.
[23] Griffiths, R.C., Milne, R.K. and R. Wood (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Austral. J. Statist. 21*(3), 238–255.
[24] Hoeffding, W. (1994). Scale-invariant correlation theory. In N. Fisher and P. Sen (Eds.), *The Collected Works of Wassily Hoeffding*, pp. 57–107. Springer, New York.
[25] Idel, M. (2016). A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps. Available at https://arxiv.org/abs/1609.06349.
[26] Ireland, C.T. and S. Kullback (1968). Contingency tables with given marginals. *Biometrika 55*(1), 179–188.
[27] Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton FL.
[28] Kateri, M. (2014). *Contingency Table Analysis: Methods and Implementation using R*. Springer, New York.
[29] Kendall, M.G. and A. Stuart (1961). *The Advanced Theory of Statistics. Vol. 2: Inference and Relationship*. Griffin, London.
[30] Kolesárová, A., Mesiar, R., Mordelová, J., and C. Sempi (2006). Discrete copulas. *IEEE Trans. Fuzzy Syst. 14*(5), 698–705.
[31] Marshall, A.W. and I. Olkin (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *J. Amer. Statist. Assoc. 80*(390), 332–338.
[32] Marshall, A.W. (1996). Copulas, marginals and joint distributions. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with Fixed Marginals and Related Topics*, pp. 213–222. Institute of Mathematical Statistics, Hayward CA.
[33] Mayor, G., Suñer, J., and J. Torrens (2005). Copula-like operations on finite settings. *IEEE Trans. Fuzzy Syst. 13*(4), 468–477.

[34] Mayor, G., Suñer, J., and J. Torrens (2007). Sklar's theorem in finite settings. *IEEE Trans. Fuzzy Syst. 15*(3), 410–416.

[35] Mikosch, T. (2006). Copulas: Tales and Facts. *Extremes 9*, 3-20.

[36] Mosteller, F. (1968). Association and Estimation in Contingency Tables. *J. Amer. Statist. Assoc. 63*(321), 1–28.

[37] Nelsen, R.B. (2006). *An Introduction to Copulas*. Second edition. Springer, New York.

[38] Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *J. Multivariate Anal. 98*(3), 544–567.

[39] Perrone, E., Solus, L., and C. Uhler (2019). Geometry of discrete copulas. *J. Multivariate Anal. 172*, 162–179.

[40] Pfeifer, D. and J. Nešlehová (2004). Modeling and generating dependent risk processes for IRM and DFA. *Astin Bull. 34*(2), 333–360.

[41] Rudas, T. (2018). *Lectures on Categorial Data Analysis*. Springer, New York.

[42] Rüschendorf, L. and W. Thomsen (1993). Note on the Schrödinger equation and *I*-projections. *Statist. Probab. Lett. 17*(5), 369–375.

[43] Rüschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. *Ann. Statist. 23*(4), 1160–1174.

[44] Rüschendorf, L. (1996). Developments on Fréchet-bounds. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with Fixed Marginals and Related Topics*, pp. 273–296. Institute of Mathematical Statistics, Hayward CA.

[45] Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

[46] Schweizer, B. and A. Sklar (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Math. 52*(1), 43–53.

[47] Schweizer, B. and E. Wolff (1981). On nonparametric measures of dependence for random variables. *Ann. Statist. 9*(4), 879–885.

[48] Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist. 35*(2), 876–879.

[49] Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly 74*(4), 402–405.

[50] Sklar, M. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 299–331.

[51] Tajar, A., Denuit, M. and Lambert, P. (2001). Copula-type representation for random couples with Bernoulli margins. Available at http://hdl.handle.net/2078.1/115376.

[52] Tankov, P. (2015). Lévy copulas: review of recent results. In M. Podolskij, R. Stelzer, S. Thorbjørnsen, and A. E. D. Veraart (Eds.), *The Fascination of Probability, Statistics and their Applications*, pp. 127–150. Springer, Cham.

[53] Yule, G.U. (1912). On the methods of measuring association between two attributes. *J. Roy. Statist. Soc. 75*(6), 579–652.

[54] Zilko, A.A. and D. Kurowicka (2016). Copula in a multivariate mixed discrete-continuous model. *Comput. Statist. Data Anal. 103*, 28–55.