Characterization of multi-way binary tables with uniform margins and fixed correlations

Roberto Fontana, Elisa Perrone, and **Fabio Rapallo**

Department of Economics
Università di Genova, Italy

ENBIS-26 Conference
Firenze, 6–10 September 2026

Dependence modeling and copulas

Sklar's Theorem (1959)

Given $\mathbf{F}_\mathbf{X}$ a joint distribution function of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with marginals $(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))$, there exists a copula C such that for all $\mathbf{x} = (x_1, \ldots, x_d)$, it follows:

$$\mathbf{F}_\mathbf{X}(\mathbf{x}) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)). \tag{1}$$

C is uniquely defined on $\mathrm{Ran}(F_{X_1}) \times \ldots \times \mathrm{Ran}(F_{X_d})$.

1. *Copulas* are tools to describe the dependence between random variables disregarding the marginal effect.
2. **Question:** Is there a natural analogous of the concept of *copulas* in the discrete setting (e.g., for Bernoulli distributions)?

A new perspective that borrows ideas from the contingency table framework

DE GRUYTER Depend. Model. 2020; 8:417–440

Research Article Open Access

Gery Geenens*

Copula modeling for discrete random vectors

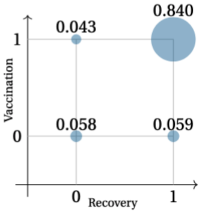
https://doi.org/10.1515/demo-2020-0022
Received September 11, 2020; accepted November 30, 2020

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.

Key ingredients of copula modeling in the discrete framework:

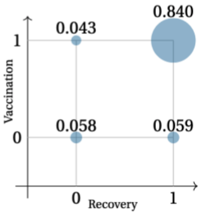
1. margin transformation to uniform distributions
2. (unique) identification of the dependence model

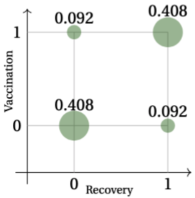
Classical example (Yule 1912)



X_1	X_2	N
Vaccinated	Recoveries	3951
Unvaccinated	Recoveries	278
Vaccinated	Deaths	200
Unvaccinated	Deaths	274

Table 1: Sheffield smallpox





Odds Ratio

$$\omega_2(X_1, X_2) = \frac{p_{00}p_{11}}{p_{01}p_{10}} = 19.4708$$

Moving to higher dimensions ($d > 2$)

Two possible approaches are considered here:

1. *Conditional odds ratios.* For each pair (X_i, X_j), we may condition on a value of the remaining variables.
2. *Marginal two-way odds ratios.* We may consider the odds ratios of the 2×2 tables obtained by summing over (collapsing) the remaining variables.

Different ways to interpret association

1. *(Higher-order) Conditional odds ratios.*
   The problem becomes: find a table with same odds ratio structure and uniform margins.
2. *Marginal two-way odds ratios.* (These do not identify a single table, but they relate to pairwise correlations when margins are fixed.)
   The problem becomes: generate tables with prescribed pairwise correlations (or marginal two-way odds ratios) and uniform margins.

Our related research

1. *Conditional odds ratios.* **Problem statement 1:** Given a table T, find a table $\tilde{T}$ with uniform margins and the same **conditional** odds ratios structure as T.

   R. Fontana, E. Perrone, F. Rapallo. **Zero patterns in multi-way binary contingency tables with uniform margins**, International Journal of Approximate Reasoning (2026)

2. *Marginal odds ratios.* **Problem statement 2:** Given d binary random variables with a prescribed pairwise dependence structure (expressed through correlations or marginal odds ratios), determine the full set of joint distributions with uniform margins that satisfy these constraints.

   R. Fontana, E. Perrone, F. Rapallo. **Characterization of multi-way binary tables with uniform margins and fixed correlations**, Statistical Papers (2026, in press), arXiv:2601.07369

What do these papers have in common?

IDEA: Fixing the margins to uniform introduces symmetries.

We exploit these symmetries and use a geometric approach to define the associated space.

1. The **first problem** is about finding which zero-patterns are compatible with a table with uniform margins (so when the associated space is non-empty).
2. The **second problem** is about defining the full convex space of joint distributions with uniform margins and dependence constraints.

I focus on the second problem here.

Notation

- d binary r.v.'s, $X_1, \ldots, X_d$, i.e., a d-dim r. v. $(X_1, \ldots, X_d)$
- k-th cell of a table corresponds to $\alpha_k = (\alpha_{k_1}, \ldots, \alpha_{k_d})$ s.t. $\sum_{j=1}^{d} \alpha_{k_j} 2^{d-j} = k - 1, \ k = 1, \ldots, 2^d$
- the table is given by the 2^d-vector of probabilities $p = (p_\alpha, \alpha \in \{0,1\}^d)$
- p belongs to the **closed simplex**

$$\Delta = \left\{ p_\alpha, \alpha \in \{0,1\}^d, p_\alpha \geq 0, \sum_{\alpha} p_\alpha = 1 \right\}$$

- $E[X_i]$ denoted by μ_i, and $E[X_i X_j]$ by μ_{ij}
- For $\mathcal{D} = \{0,1\}^d$, d margins $m_i, i = 1, \ldots, d$ defined as

$$m_i = (m_i^0, m_i^1) = \left(\sum_{\alpha \in \mathcal{D}, \alpha_i = 0} p_\alpha, \sum_{\alpha \in \mathcal{D}, \alpha_i = 1} p_\alpha \right).$$

Notation

- Consider the $\binom{d}{2}$ bivariate margins

$$m_{ij} = \begin{pmatrix} m_{ij}^{00} & m_{ij}^{01} \\ m_{ij}^{10} & m_{ij}^{11} \end{pmatrix},$$

where

$$m_{ij}^{k_1 k_2} = \sum_{\alpha \in \mathcal{D}, \alpha_i = k_1, \alpha_j = k_2,} p_\alpha,$$

$k_1, k_2 \in \{0, 1\}$, and $i, j = 1, \ldots, d, i < j$

Note that $m_{ij}^{11} = \mu_{ij}$, and if the margins m_i and m_j are fixed to $\overline{m}_i = (\overline{m}_i^0, \overline{m}_i^1)$ and $\overline{m}_j = (\overline{m}_j^0, \overline{m}_j^1)$, m_{ij} becomes a function of m_{ij}^{11}:

$$m_{ij} = \begin{pmatrix} \overline{m}_i^0 - \overline{m}_j^1 + m_{ij}^{11} & \overline{m}_j^1 - m_{ij}^{11} \\ \overline{m}_j^1 - m_{ij}^{11} & m_{ij}^{11} \end{pmatrix}$$

Conditional versus marginal odds ratios

1. For any pair (X_i, X_j), we define conditional odds ratios by fixing the values of the remaining $(d-2)$ variables.
2. The marginal odds ratio for a pair (X_i, X_j) is obtained by summing over all other variables:

$$\omega_M^{ij} = \frac{\sum_{\alpha:\alpha_i=\alpha_j=1} p_\alpha \ \sum_{\alpha:\alpha_i=\alpha_j=0} p_\alpha}{\sum_{\alpha:\alpha_i=1,\alpha_j=0} p_\alpha \ \sum_{\alpha:\alpha_i=0,\alpha_j=1} p_\alpha} = \frac{m_{ij}^{11} m_{ij}^{00}}{m_{ij}^{10} m_{ij}^{01}} . \tag{2}$$

Remarks from Algebraic Statistics

1. Conditional odds ratios relate to classical log-linear models, and they define toric varieties (defined through binomial equations): MCMC with Markov bases is possible.
2. Marginal odds reations relate to marginal log-linear models, they don't define toric varieties: we need to adopt a different combinatorial description.

Problem statement with uniform margins

- For a table T with uniform margins:

$$m_{i0} = m_{i1} = \frac{1}{2} \Leftrightarrow m_{i0} - m_{i1} = 0 \Leftrightarrow \sum_{\alpha \in \mathcal{D}, \alpha_i = 0} p_\alpha - \sum_{\alpha \in \mathcal{D}, \alpha_i = 1} p_\alpha = 0,$$

  $(i = 1 \ldots, d)$.

- In compact form:

$$(1_{\{\alpha \in \mathcal{D}, \alpha_i = 0\}} - 1_{\{\alpha \in \mathcal{D}, \alpha_i = 1\}})^T p = 0,$$

  where 1_A is the indicator vector of A, b^T denotes the transpose of b, and p is the vector $p = (p_\alpha, \alpha \in \mathcal{D})$.

Problem statement with uniform margins

- Now, include the bivariate margins:

$$m_{ij} = \begin{pmatrix} m_{ij}^{11} & \frac{1}{2} - m_{ij}^{11} \\ \frac{1}{2} - m_{ij}^{11} & m_{ij}^{11} \end{pmatrix} . \tag{3}$$

$$\sum_{\alpha \in \mathcal{D}, \alpha_i = \alpha_j = 1} p_\alpha = \bar{\mu}_{ij},$$

that is

$$\sum_{\alpha \in \mathcal{D}, \alpha_i \alpha_j = 1} (\bar{\mu}_{ij} - 1) p_\alpha + \bar{\mu}_{ij} \sum_{\alpha \in \mathcal{D}, \alpha_i \alpha_j = 0} p_\alpha = 0. \tag{4}$$

- In compact form:

$$((\bar{\mu}_{ij} - 1) 1_{\{\alpha \in \mathcal{D}, \alpha_i \alpha_j = 1\}} + \bar{\mu}_{ij} 1_{\{\alpha \in \mathcal{D}, \alpha_i \alpha_j = 0\}})^T p = 0, \quad i, j = 1, \ldots, d, \quad i < j. \tag{5}$$

Problem formulation

With $d = 3$ putting all equations in a matrix we get:

$$H_3 p = \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ \bar{\mu}_{12} & \bar{\mu}_{12} & \bar{\mu}_{12} & \bar{\mu}_{12} & \bar{\mu}_{12} & \bar{\mu}_{12} & (\bar{\mu}_{12} - 1) & (\bar{\mu}_{12} - 1) \\ \bar{\mu}_{13} & \bar{\mu}_{13} & \bar{\mu}_{13} & \bar{\mu}_{13} & \bar{\mu}_{13} & (\bar{\mu}_{13} - 1) & \bar{\mu}_{13} & (\bar{\mu}_{13} - 1) \\ \bar{\mu}_{23} & \bar{\mu}_{23} & \bar{\mu}_{23} & (\bar{\mu}_{23} - 1) & \bar{\mu}_{23} & \bar{\mu}_{23} & \bar{\mu}_{23} & (\bar{\mu}_{23} - 1) \end{pmatrix} p$$

The polyhedral cone defined by H_d is

$$\mathcal{C}(H_d) = \{p = (p_\alpha), p_\alpha \in \mathbb{R}, p_\alpha \geq 0, \ \alpha \in \mathcal{D}, H_d p = 0, \} \tag{6}$$

and the solution set for our problem is

$$\mathcal{C}(H_d) \cap \Delta$$

Computations

We are searching for the extreme rays $\tilde{y}_{i\alpha}$ to define the space geometrically.

The extreme pmfs $r_i = (r_{i\alpha}, i = 1, \ldots, n_d, \alpha \in \mathcal{D})$ are $r_{i\alpha} = \frac{\tilde{y}_{i\alpha}}{\sum_{\alpha \in \mathcal{D}} \tilde{y}_{i\alpha}}$.

Example $d = 3$

Let us consider the table below

pmf	000	001	010	011	100	101	110	111	ω_M^{12}	ω_M^{13}	ω_M^{23}
p_0	0.1	0.05	0.3	0.2	0.1	0.05	0.15	0.05	0.40	0.64	1.11

We want to determine the convex hull of pmfs with uniform margins and prescribed pairwise correlations.

We build the H_3 matrix and find the extreme pmfs by using `4ti2`.

pmf	000	001	010	011	100	101	110	111
r_1	0	0.194	0.222	0.084	0.257	0.05	0.021	0.173
r_2	0.173	0.021	0.05	0.257	0.084	0.222	0.194	0

Remarks

Question

How does the space look like?

The space is the segment $p = \alpha r_1 + (1 - \alpha) r_2$, with $\alpha \in [0, 1]$.

$$\omega^{123} = \frac{p_{000} p_{011} p_{101} p_{110}}{p_{111} p_{100} p_{010} p_{001}}$$

tends to 0 as α approaches 1 and diverges to $+\infty$ as α approaches 0: the segment spans all possible three-way dependence structures and we can choose from there

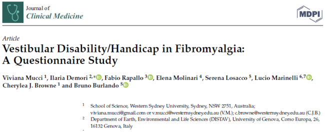
Key remarks

- The combinatorial approach works on the polyhedral cone

$$\mathcal{C}(H_d) = \{p = (p_\alpha), p_\alpha \in \mathbb{R}, p_\alpha \geq 0, \ \alpha \in \mathcal{D}, H_d p = 0, \} \ .$$

- H_d works directly on the simplex and zeros are allowed (i.e., it works on the **closed simplex**).
- The matrix H_d depends on the actual values of the margins and of the pairwise correlations.
- The extremal pmfs contain the extreme values of higher-order moments.

Example $d = 5$


Journal of Clinical Medicine
MDPI

Article

Vestibular Disability/Handicap in Fibromyalgia: A Questionnaire Study

Viviana Mucci [1], Ilaria Demori [2,*], Fabio Rapallo [3], Elena Molinari [4], Serena Losacco [5], Lucio Marinelli [6,7], Cherylea J. Browne [1] and Bruno Burlando [5]

1 School of Science, Western Sydney University, Sydney, NSW 2751, Australia; viviana.mucci@gmail.com or v.mucci@westernsydney.edu.au (V.M.); c.browne@westernsydney.edu.au (C.J.B.)
2 Department of Earth, Environmental and Life Sciences (DISTAV), University of Genova, Corso Europa, 26, 16132 Genova, Italy

In a study about symptoms in patients with fibromyalgia, five typical symptoms are registered (1 if present, 0 if absent).

- Photophobia
- Migraine
- Brain fog
- Nausea
- Dizziness

We use information about pairwise correlation to describe all probability distributions with uniform margins and prescribed pairwise correlations.

Example $d = 5$

With the matrix H_5 and using `4ti2` we obtain $181,588$ extreme pmfs which characterize the polytope. Moreover, such extreme pmfs contain all extreme values of higher order interactions:

x_1	x_2	x_3	x_4	x_5	min	max
0	0	1	1	1	0.119	0.297
0	1	0	1	1	0.139	0.313
0	1	1	0	1	0.121	0.258
0	1	1	1	0	0.121	0.258
0	1	1	1	1	0	0.258
1	0	0	1	1	0.112	0.285
1	0	1	0	1	0.13	0.299
1	0	1	1	0	0.12	0.285
1	0	1	1	1	0	0.274
1	1	0	0	1	0.114	0.295
1	1	0	1	0	0.119	0.285
1	1	0	1	1	0	0.277
1	1	1	0	0	0.118	0.258
1	1	1	0	1	0	0.258
1	1	1	1	0	0	0.258
1	1	1	1	1	0	0.253

Take home message and future work

In this talk:

- we showed the practical advantages of working directly on the simplex (and restricting to the case of uniform margins) for contingency table analysis;
- we provided an elegant geometric approach to solve practical questions on table generation with fixed margins and prescribed correlations.

Moving forward:

- investigate how the zero-pattern (i.e., structural zeros) would change/restrict the admissible space;
- improve the available sampling techniques to obtain elements from the derived convex space.

References

- R. Fontana, E. Perrone, F. Rapallo. **Zero patterns in multi-way binary contingency tables with uniform margins**, International Journal of Approximate Reasoning (2026)
- Gery Geenens. **Copula modeling for discrete random vectors**, Dependence Modeling, vol 8, nr.1, pp. 417–440 (2020).
- Gery Geenens, Ivan Kojadinovic, Tommaso Martini. **The empirical discrete copula process**, arXiv:2506.12316v1 (2025).
- Ivan Kojadinovic and Tommaso Martini. **Copula-like inference for discrete bivariate distributions with rectangular supports**, Electronic Journal of Statistics 18 (1), 2571–2619 (2024)
- Tamás Rudas. **Lectures on categorical data analysis**. *Springer*, 2018.

*R. Fontana, E. Perrone, F. Rapallo. **Characterization of multi-way binary tables with uniform margins and fixed correlations**, Statistical Papers, in press, arXiv:2601.07369*

Thank you! Any questions?