arXiv:2601.07369v1 [stat.ME] 12 Jan 2026

具有均匀边际和固定相关性的多维二元列联表的刻画

Roberto Fontana[1]，Elisa Perrone[2]，Fabio Rapallo[3*]

[1]数学科学系，都灵理工大学，Corso Duca degli Abruzzi 24，都灵，10129，意大利。[2]数学与计算机科学系，埃因霍温理工大学，Groene Loper 3，埃因霍温，5612 AE，荷兰。[3*]经济学系，热那亚大学，via Vivaldi 5，热那亚，16126，意大利。

通讯作者。电子邮箱：fabio.rapallo@unige.it；贡献作者：roberto.fontana@polito.it；e.perrone@tue.nl；

摘要

在许多涉及二元变量的应用中，我们只能获得相关性等两两依赖度量。然而，对于包含两个以上变量的多维列联表，这些量无法唯一确定联合分布，而是定义了一族具有相同两两依赖、但高阶交互可能不同的可容许分布。本文引入一个几何框架，用于描述具有均匀边际的此类联合分布的完整可行集。我们证明，该可容许集构成一个凸多胞体，并分析其对称性，刻画其极射线。这些极端分布有助于深入理解：在保持给定两两信息的同时，高阶依赖结构可能如何变化。与传统的列联表生成方法不同，后者只返回一个列联表；我们的框架能够探索并理解依赖结构的整个可容许空间，从而为建模和模拟提供更灵活的选择。我们通过实例以及一个关于评分者一致性的真实案例研究，展示了理论结果的实用价值。

关键词：列联表、离散Copula、优势比、评分者一致性

1 引言

对分类变量和二元变量之间依赖结构进行建模，是统计学中的核心问题，也是分析列联表的关键，在模拟、推断和模型选择中有重要应用（例如参见 [1, 2]）。研究多元依赖在评分者一致性分析等应用中尤为重要，因为关联性度量在这些情境下尤其难以处理（例如参见 [3]）。对于两个分类变量或二元变量，情况相对明晰：当边际分布固定时，相关性与优势比相关，并能唯一确定联合分布 [1]。然而，维数升高后，情况便大为不同。当变量数量增至三个或更多时，相关系数（或边际优势比）不再能唯一确定联合分布 [1], [4]。相反，它们确定了一族可行分布，这些分布具有相同的成对依赖结构，但高阶交互作用各不相同。

在实际应用中，通常使用基于对数线性模型 [5] 或多元迭代比例拟合程序（MIPFP） [6] 的方法，生成具有指定边际分布和两两相关性的概率列联表。然而，这些方法本质上会从与给定两两相关性相容的可行联合分布集合中选取一个分布。此类选择往往未明确承认可接受解并非唯一，也未阐明所选建模策略如何隐式地确定高阶依赖。因此，这些方法可能会施加非预期的结构约束。在评分者一致性分析等应用中，这一局限尤为突出：MIPFP 往往会将高阶优势比强制设为 1，而对数线性模型则会施加特定的层次依赖结构，从而可能导致对联合一致性的表示不切实际或限制过严。

本文旨在刻画具有固定两两依赖关系（通过相关系数或边际优势比表示）和均匀边际的联合分布的整个可行集。本研究补充了近期关于列联表中的优势比和零模式相容性的研究 [7], ，将重点转向固定两两相关系数所形成的几何结构。通过关注完整的可容许集，而非由单一模型生成的表，我们将依赖建模与对数线性表示联系起来，提供了一个框架，使人们能够在可容许族中更审慎、更明智地选择概率表。我们采用几何视角：将可容许分布的可行区域视为凸多胞体，其顶点和结构编码了所有符合给定约束的可能联合分布。将研究范围限定为具有均匀边际的分布，灵感来自离散随机向量的 copula 理论近期进展 [8–13], ，其核心思想是描绘依赖关系，同时不考虑边际行为。将列联表限定为具有均匀边际，对我们的方法有明显优势：通过聚焦于均匀边际，我们得以获得依赖结构的清晰几何表示，并揭示联合分布空间中优雅的对称性；这些对称性对于描述和刻画整个可行集至关重要。这种几何刻画对于以下应用尤其相关：已知或需要两两依赖信息，但

高阶交互作用尚不确定或具有灵活性。例如，评定者一致性分析中，超越两两层面的联合一致性的真实性质可能有所不同；又如模拟和缺失数据插补任务，探索多个可行的概率表既有意义，也属必要。事实上，在评定者一致性分析中，众所周知，边际非齐性会使一致性度量难以解释 [14]，因此文献中已经探索过的一种解决方案 [15] 是将原始表转换为具有齐性边际（即均匀边际）的等价表。通过刻画具有均匀边际的二元列联表的整个可容许类别，我们的框架为二元变量之间的依赖结构提供了更具信息量且更实用的视角。尽管本文聚焦于二元列联表，我们也认为将该理论扩展到多维非二元列联表，是未来研究的一个重要方向。

本文其余部分安排如下。第2 节介绍符号表示、问题表述、与对数线性模型的联系，以及生成具有指定相关系数的列联表的现有方法。第3 节给出可行集的几何刻画，包括其多面体结构。第4 节通过示例展示该方法，并说明其在实际评定者一致性研究中的适用性。最后，第5 节总结全文，并探讨未来研究的可能方向。

2 背景与相关工作

本节首先介绍符号表示，并回顾用于描述二元变量之间依赖关系的关键概念。具体而言，我们回顾两两相关性和优势比的概念，讨论二者之间的关系，并强调它们在确定联合分布依赖结构中的作用。

我们考虑 $d \geq 2$ 二元随机变量 X_1，……，X_d，其联合单元概率为

$$p_\alpha = \mathbb{P}(X_1 = \alpha_1, \ldots, X_d = \alpha_d), \qquad \alpha = (\alpha_1, \ldots, \alpha_d) \in \{0,1\}^d.$$

必要时，我们假设集合 $\{0,1\}^d$ 按字典序排列，由此可以指称 k的第 T、 $k = 1$、……、2^d 个单元。我们注意到，列联表的第 k 个单元对应于 $\alpha_k = ((\alpha_k)1$，……，$(\alpha_k)_d)$，其中 $\sum_{j=1}^{d}(\alpha_k)_j 2^{d-j} = k-1$， $k =1$，……，2^d；也就是说， α_k 是 $k-1$的二进制表示。因此，列联表 T由 2^d维向量 $p = (p_\alpha$， $\alpha \in \{0,1\}^d)$描述。为简化符号表示，我们将 $\{0,1\}^d$ 记为$\mathcal{D}$，将 $E[X_i]$ 记为 μ_i，并将 $E[X_iX_j]$ 记为 μ_{ij}。给定 $\alpha = (\alpha_1$，……，$\alpha_d)$，我们用 $1-\alpha$ 表示向量（$1-\alpha_1$，……，$1-\alpha_d$）。

有如下定义的 d 单变量边际分布 m_i、 $i = 1$、……、 d

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

我们观察到 $m_i^1 = \mu_i$ 和 $m_i^0 = 1 - \mu_i$, $i = 1,\ldots, d$。

我们还考虑 $\binom{d}{2}$ 二元边缘分布

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

其中

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

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

我们注意到 $m_{ij}^{11} = \mu_{ij}$，并且如果将单变量边际分布 m_i 和 m_j 分别固定为 $\overline{m}_i = (\overline{m}_i^0, \overline{m}_i^1)$ 和 $\overline{m}_j = (\overline{m}_j^0, \overline{m}_j^1)$，就可以将 m_{ij} 表示为 m_{ij}^{11} 的函数：

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

对于任意 $i \neq j \in \{1, \ldots, d\}$，$X_i$ 与 X_j 之间的成对相关仅取决于边缘概率以及对所有其他变量求和后得到的边际 2×2 子表。它可以表示为

$$\begin{aligned} \mathrm{Cor}(X_i, X_j) &= \frac{\mathbb{P}(X_i = 1, X_j = 1) - \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1)}{\sqrt{\mathbb{P}(X_i = 1)(1 - \mathbb{P}(X_i = 1))\,\mathbb{P}(X_j = 1)(1 - \mathbb{P}(X_j = 1))}} \\ &= \frac{m_{ij}^{11} - m_i^1 m_j^1}{\sqrt{m_i^1 m_i^0 m_j^1 m_j^0}}. \end{aligned} \tag{2}$$

对于任意一对 (X_i, X_j)，我们通过固定其余 $(d-2)$ 个变量的取值来定义条件优势比。令 $\alpha' \in \{0,1\}^{d-2}$ 表示其他变量的配置。那么，在给定 $X_{-(i,j)} = \alpha'$ 时，(X_i, X_j) 的条件优势比为

$$\omega_{\alpha'}^{ij} = \frac{p_{\alpha_1}\, p_{\alpha_2}}{p_{\alpha_3}\, p_{\alpha_4}}, \tag{3}$$

其中，四个索引向量 $\alpha_1, \ldots, \alpha_4 \in \{0,1\}^d$ 在 α' 外部位置 i 和 j 上的取值相同，而它们的 (i, j) 项分别为：

$$\alpha_1(i,j) = (1,1), \quad \alpha_2(i,j) = (0,0), \quad \alpha_3(i,j) = (1,0), \quad \alpha_4(i,j) = (0,1).$$

例如，在 $d = 4$ 中，固定 $(X_2, X_4) = (0,1)$ 可得：

$$\omega_{01}^{13} = \frac{p_{0001}\, p_{1011}}{p_{0011}\, p_{1001}}.$$

在一个 d 维二元表中，有 $\binom{d}{2} 2^{d-2}$ 个这样的条件优势比，不过它们并非全部相互独立。

一对变量（X_i， X_j）的边际优势比，可通过对所有其他变量求和得到：

$$\omega_M^{ij} = \frac{\sum_{\alpha:\alpha_i=\alpha_j=1} p_\alpha \sum_{\alpha:\alpha_i=\alpha_j=0} p_\alpha}{\sum_{\alpha:\alpha_i=1,\alpha_j=0} p_\alpha \sum_{\alpha:\alpha_i=0,\alpha_j=1} p_\alpha} = \frac{m_{ij}^{11} m_{ij}^{00}}{m_{ij}^{10} m_{ij}^{01}}. \tag{4}$$

在这种情形下，变量之间的依赖结构可以用不同方式概括，最常见的是通过两两相关性，或通过边际和条件优势比。在二元情形中，这些概念彼此一致：给定边际后，无论采用哪一种，都能唯一确定联合分布；但在高维情形中，它们则有所不同，只能约束一族可接受的分布。在下文中，我们将分别考察二元和多元情形，以凸显这些差异，并讨论相关性和优势比如何确定（或无法确定）潜在的联合分布。在此过程中，我们将说明不同的优势比结构如何通过其 copula 表示与完整联合分布相关联，并自然地联系到本研究所探讨的主要问题 [9, 13, 16]。这一视角将为我们随后对具有固定两两依赖和均匀边际的分布可行集进行几何分析奠定概念基础。

2.1 双向二元列联表

我们从简单的二元 2×2 列联表情形开始。在这种情形下， X_1 与 X_2 之间的依赖关系既可以用相关性描述，也可以用优势比描述，而这两种度量密切相关。

X_1 与 X_2 之间的皮尔逊相关系数可写为

$$\begin{aligned}
\mathrm{Cor}(X_1, X_2) &= \frac{\mathbb{P}(X_1 = 1, X_2 = 1) - \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)}{\sqrt{\mathbb{P}(X_1 = 1)(1 - \mathbb{P}(X_1 = 1))\mathbb{P}(X_2 = 1)(1 - \mathbb{P}(X_2 = 1))}} \\
&= \frac{p_{11} - m_1^1 m_2^1}{\sqrt{m_1^1(1 - m_1^1)\, m_2^1(1 - m_2^1)}} \\
&= \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{m_1^1 m_1^0 m_2^1 m_2^0}}.
\end{aligned} \tag{5}$$

优势比也是一种可行的方式，可用于描述2×2 列联表中的依赖关系。具体而言，在一个 2×2 列联表中，只有一个优势比，其定义为

$$\omega^{12} = \frac{p_{11}p_{00}}{p_{10}p_{01}}.$$

相关性的分子，即协方差 $\mathrm{Cov}(X_1, X_2)$，与$p_{11}p_{00} - p_{10}p_{01}$ 一致，并且与优势比直接相关：当且仅当 X_1 和 X_2 相互独立时， $\omega^{12} = 1$；在这种情况下，$\mathrm{Cor}(X_1, X_2) = 0$。更一般地， $\omega^{12} > 1$ 意味着正关联，而 $\omega^{12} < 1$ 意味着负关联[1, 2]。

均匀边际下通过优势比体现的依赖关系。将表格的边际转换为均匀边际后，所得表格由优势比唯一确定

（或等价地，由相关性唯一确定），这与近期 Copula 理论的研究结果一致。 [9]事实上，在二元二值情形下，存在唯一的概率分布 p具有均匀边际，即

$$\overline{m}_1^1 = \overline{m}_2^1 = \frac{1}{2},$$

使得 p保留原有的优势比 ω^{12}。这意味着，在边际固定为均匀边际后，优势比（等价地，相关性）便能完全刻画依赖结构。换言之，在维数为 $d = 2$ 时，具有固定两两依赖和均匀边际的分布所构成的可行集缩减为单个点。这个点还受到沿对角线对称的约束，即 $p_{00} = p_{11}$ 且 $p_{01} = p_{11} = \frac{1}{2} - p_{00} = \frac{1}{2} - p_{11}$。此外，在这种情况下，通过对方程（5）作简单代换，我们可以将 $\mathrm{Cor}(X_1, X_2)$ 重写为关于 p_{11} 的线性函数，即 $\mathrm{Cor}(X_1, X_2) = 4p_{11} - 1$。

对于 $d \geq 3$，解集并非平凡，但仍保留了二元情形中观察到的对称性。这凸显了二元框架与多元框架之间的主要概念差异；在高维情况下，这也说明了我们为何强调描述完整的可容许集，而非只关注某个具有代表性的分布。

2.2 多维二元列联表

当 $d \geq 3$ 时，相关系数与优势比之间的关系会变得更加复杂，仅固定所有两两相关性（或等价地，所有边际优势比）已不足以唯一确定联合分布。

为说明这种情况，我们来考虑 $d = 3$，即一个随机向量（X_1，X_2，X_3）。根据式（2），X_1 与 X_2 之间的成对相关为

$$\mathrm{Cor}(X_1, X_2) = \frac{m_{12}^{11} m_{12}^{00} - m_{12}^{10} m_{12}^{01}}{\sqrt{m_1^1 m_1^0 m_2^1 m_2^0}}.$$

$\mathrm{Cor}(X_1, X_3)$ 和 $\mathrm{Cor}(X_2, X_3)$ 也有类似的公式。

条件优势比与边缘优势比。在 $d = 3$ 的情况下，存在两种本质上不同的优势比。

(*i*) 条件优势比。对于每一对（X_i，X_j），我们可以对剩余变量的某个取值进行条件化。根据式（3），可得到六个条件优势比：

$$\begin{aligned}
\omega_0^{23} &= \frac{p_{000}p_{011}}{p_{001}p_{010}}, & \omega_1^{23} &= \frac{p_{100}p_{111}}{p_{101}p_{110}}, \\
\omega_0^{13} &= \frac{p_{000}p_{101}}{p_{001}p_{100}}, & \omega_1^{13} &= \frac{p_{010}p_{111}}{p_{011}p_{110}}, \\
\omega_0^{12} &= \frac{p_{000}p_{110}}{p_{010}p_{100}}, & \omega_1^{12} &= \frac{p_{001}p_{111}}{p_{011}p_{101}}.
\end{aligned} \tag{6}$$

这六个约束中只有四个是独立的。

(*ii*) 边际优势比。我们也可以考虑以下 2×2 列联表的优势比：根据公式（4），对其余变量求和（合并）得到：

$$\omega_M^{12} = \frac{m_{12}^{11} m_{12}^{00}}{m_{12}^{10} m_{12}^{01}}, \qquad \omega_M^{23} = \frac{m_{23}^{11} m_{23}^{00}}{m_{23}^{10} m_{23}^{01}}, \qquad \omega_M^{13} = \frac{m_{13}^{11} m_{13}^{00}}{m_{13}^{10} m_{13}^{01}}. \tag{7}$$

条件优势比与边际优势比通常并不一致。一对变量（X_i，X_j）之间的所有条件优势比都可能等于 1（这表明条件独立），而边际优势比 ω_M^{ij} 却显示出很强的关联。这一经典现象称为辛普森悖论，说明边际优势比描述的是合并列联表中的关联，而条件优势比反映的是在给定其他变量条件下的依赖关系 [1, 2]。

相关性、边际优势比与依赖关系。两个二元随机变量 X_i 与 X_j 之间的相关性与相应的边际优势比 ω_M^{ij} 直接相关，但与条件优势比无关。因此，即使边际均匀，指定所有两两相关性（或等价地，所有边际优势比）也无法确定条件优势比或任何高阶交互项。这一点至关重要：在任意维数下，条件优势比和均匀边际可以唯一确定完整的联合分布（例如，见[9, 13, 16]及其中的参考文献），但边际优势比（即成对相关）则不能。因此，固定所有成对相关和边际并不能确定唯一的联合分布，而是确定一个分布族。本节第 3 节将重点分析这一可行分布族的结构并对其进行几何刻画。本节余下部分将简要介绍从相关二元分布中抽样的主要算法，以强调我们几何描述的重要性。

2.3 相关二元随机变量的抽样算法

在模拟研究中，生成具有给定相关结构的多元二元随机变量是一项关键任务，文献中已提出各种各样的方法来解决这一问题。随机生成在推断和模型验证等任务中发挥着重要作用。本文回顾与我们的理论密切相关的几种主要方法。

文献 [17] 提出了一种线性规划方法，通过对单元概率施加线性约束或最小化某些线性函数来构造联合分布。该方法可使用标准线性规划求解器轻松实现。然而，随着列联表维数的增加，所需的大量线性约束可能会使得到的分布具有相当大的任意性。

在 [6], 中，作者提出了一种通用方法，可将一个多维列联表转换为另一个具有指定边际、同时保留原始优势比结构的多维列联表。由于其二维对应方法——迭代比例

拟合是统计学和数值分析中的经典方法 [18]，这一多维扩展称为 MIPFP。作为 MIPFP 的副产品，[6] 作者还提出了一种构造具有指定成对边际优势比的多元伯努利分布的方法

$$\omega_M^{ij} = \frac{\left(\sum_{\alpha_i=1,\alpha_j=1} p_\alpha\right)\left(\sum_{\alpha_i=0,\alpha_j=0} p_\alpha\right)}{\left(\sum_{\alpha_i=0,\alpha_j=1} p_\alpha\right)\left(\sum_{\alpha_i=1,\alpha_j=0} p_\alpha\right)} = \frac{m_{ij}^{11} m_{ij}^{00}}{m_{ij}^{10} m_{ij}^{01}},$$

并同时将列联表中的所有高阶优势比设为零。该过程在R 中通过`mipfp` 包的`ObtainMultBinaryDist`函数实现。此方法计算效率高，只需简单应用 MIPFP 即可得到所需结果。该算法适用于任意大小的列联表，不过尤其适合处理二元表。

最后，在 [5], 中，作者利用联合分布的对数线性形式生成列联表。相关性用优势比表示，并通过采用角参数化以及适当的代数化简步骤，得到有效的联合分布。例如，对于一个三维二元表，若假设不存在三阶交互项，则模型可表示为

$$\log p_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{12} + \lambda_{jk}^{13}.$$

将该模型用优势比重新参数化后，可写为

$$\log p_{ijk} = \lambda + \lambda_i^1 + \lambda_j^1 + \lambda_k^3 + \omega_M^{12} + \omega_M^{13} + \omega_M^{23}.$$

该方法的主要优势在于，它可以推广到具有任意水平数的多水平列联表。不过，随着维数增加，相关系数与优势比之间的联系可能需要复杂的方程，这使得难以根据边际优势比选择参数。同一篇论文 [5], 的引言部分简要回顾了其他基于对数线性模型理论的方法。

上述讨论的所有方法都依赖于选择一个单一的联合分布，通常还会施加超出给定成对依赖关系的额外假设。正如引言中所强调的，我们的目标则是描述所有可行分布构成的完整集合，从而系统地呈现所有可实现的依赖结构；这将是下一节的主题。

3 具有均匀边际和固定相关性的多维二元列联表的几何刻画

现在，我们将理论发展的核心问题形式化。问题陈述：给定 d 个具有预设成对依赖结构（二元随机变量之间的相关系数或边际优势比）的二元随机变量，

确定满足这些约束且具有均匀边际的联合分布的完整集合。具体而言，描述这一可容许集的几何结构，理解其中高阶交互作用如何变化，并确定其极值元素。

如第 2 节所述，将讨论范围限定在具有均匀边际的列联表上，在数学和统计上都有优势。从数学角度看，均匀边际会使列联表具有对称性，从而更清晰地几何刻画可行集。从统计角度看，采用均匀边际可以消除边际分布不均衡的影响，更直接地控制依赖结构。为了在我们的框架中纳入均匀边际，下面以代数形式表述均匀边际条件。具体而言，要获得均匀边际，列联表 p T 的概率质量函数（pmf）必须满足以下约束

$$m_i^0 = m_i^1 = \frac{1}{2} \Leftrightarrow m_i^0 - m_i^1 = 0 \Leftrightarrow \sum_{\alpha\in\mathcal{D},\alpha_i=0} p_\alpha - \sum_{\alpha\in\mathcal{D},\alpha_i=1} p_\alpha = 0,\ i = 1,\ldots,d,$$

可改写为

$$(1_{\{\alpha\in\mathcal{D},\alpha_i=0\}} - 1_{\{\alpha\in\mathcal{D},\alpha_i=1\}})^T p = 0,\ \ i = 1,\ldots,d, \tag{8}$$

其中， 1_A 是 A的指示向量， b^T 表示 b的转置， p 是向量$p = (p_\alpha,\ \alpha \in \mathcal{D})$。

如果除均匀边际外，我们还想指定二元边缘分布，则可见，式（1）变为

$$m_{ij} = \begin{pmatrix} m_{ij}^{11} & \frac{1}{2} - m_{ij}^{11} \\ \frac{1}{2} - m_{ij}^{11} & m_{ij}^{11} \end{pmatrix}. \tag{9}$$

于是，要指定均匀边际，只需指定 $m_{ij}^{11} = m_{ij}^{11}$ ，或等价地指定 $\mu_{ij} = \mu_{ij}$、$\overline{i}$, $j = 1$、.....、 d, $i \overline{<} j$。我们得到

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

也就是说

$$\sum_{\alpha\in\mathcal{D},\alpha_i\alpha_j=1} (\overline{\mu}_{ij} - 1)p_\alpha + \overline{\mu}_{ij} \sum_{\alpha\in\mathcal{D},\alpha_i\alpha_j=0} p_\alpha = 0. \tag{10}$$

我们可以将公式（10）写成

$$((\overline{\mu}_{ij} - 1)1_{\{\alpha\in\mathcal{D},\alpha_i\alpha_j=1\}} + \overline{\mu}_{ij}1_{\{\alpha\in\mathcal{D},\alpha_i\alpha_j=0\}})^T p = 0,\ \ i,j = 1,\ldots,d,\ \ i < j. \tag{11}$$

方程（8）和方程（11）可以通过将其记为 H_d 系数矩阵的方式写成紧凑形式。即，方程（8）和方程（11）变为:

$$H_d\, p = 0\,. \tag{12}$$

因此，由 H_d 定义的多面锥为

$$\mathcal{C}(H_d) = \{p = (p_\alpha), p_\alpha \in \mathbb{R}, p_\alpha \geq 0,\ \alpha \in \mathcal{D}, H_d p = 0, \}\ . \tag{13}$$

例如，对于 $d = 3$ 我们得到:

$$H_3 p = \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ \overline{\mu}_{12} & \overline{\mu}_{12} & \overline{\mu}_{12} & \overline{\mu}_{12} & \overline{\mu}_{12} & \overline{\mu}_{12} & (\overline{\mu}_{12}-1) & (\overline{\mu}_{12}-1) \\ \overline{\mu}_{13} & \overline{\mu}_{13} & \overline{\mu}_{13} & \overline{\mu}_{13} & \overline{\mu}_{13} & (\overline{\mu}_{13}-1) & \overline{\mu}_{13} & (\overline{\mu}_{13}-1) \\ \overline{\mu}_{23} & \overline{\mu}_{23} & \overline{\mu}_{23} & (\overline{\mu}_{23}-1) & \overline{\mu}_{23} & \overline{\mu}_{23} & \overline{\mu}_{23} & (\overline{\mu}_{23}-1) \end{pmatrix} p = 0\,.$$

齐次系统 $H_d p = 0$ 在满足 $p_\alpha \geq 0$ 时的解构成多面锥 $\mathcal{C}(H_d)$，其定义见公式（13）。该锥的一个极射线是一个非零向量 $\tilde{y} \in \mathcal{C}(H_d)$，它不能表示为该锥中其他元素的非平凡线性组合。因此，极射线代表了该锥的基本生成方向：每个元素 $p \in \mathcal{C}(H_d)$都可以写成其极射线的非负线性组合。因此，确定这些生成射线对于理解可行集的结构至关重要。系统 $H_d y = 0$的极射线 $\tilde{y}_i = (\tilde{y}_{i\alpha},\ i = 1, \ldots, n_d,\ \alpha \in \mathcal{D}, \tilde{y}_\alpha \geq 0)$，其中$y = (y_\alpha, \alpha \in \mathcal{D}, y_\alpha \geq 0)$，与公式（13）中引入的多面锥 $\mathcal{C}(H_d)$的极射线重合（见 [19, 20]）。可以使用交换代数软件计算这些极射线，例如本工作中使用的 4ti2 [21]，。随后，通过对 $r_i = (r_{i\alpha}, i = 1, \ldots, n_d, \alpha \in \mathcal{D})$进行简单归一化 $r_{i\alpha} = \frac{y_{i\alpha}}{\sum_{\alpha \in \mathcal{D}} \tilde{y}_{i\alpha}}, i = 1, \ldots, n_d, \alpha \in \mathcal{D}$，即可计算出相应的极端概率质量函数。多面锥 $\mathcal{C}(H_d)$与单纯形的交集

$$\Delta = \left\{ p_\alpha \in \mathbb{R}^{(2^d)} \ :\ p_\alpha \geq 0,\ \sum_\alpha p_\alpha = 1 \right\}$$

是凸多胞体 $\mathcal{F}(H_d)$，其极点为极端概率质量函数 $r_i,\ i$ =1、.....、 n_d:

$$\mathcal{F}(H_d) = \left\{ p = (p_\alpha), p_\alpha \in \mathbb{R}, p_\alpha \geq 0,\ \alpha \in \mathcal{D}, H_d p = 0, \sum_{\alpha \in \mathcal{D}} p_\alpha = 1 \right\}\ . \tag{14}$$

值得注意的是，给定极端概率质量函数 $r_i,\ i = 1,\ \ldots,\ n_d$其中 $r_i,\ i$ =1，...，n_d和 n_d 取决于维数 d以及多维列联表指定的单变量和二元边缘分布，任何满足公式（12）的概率质量函数 p 都可以写成

$$p = \sum_{i=1}^{n_d} \theta_i r_i$$

其中 $\theta_i \geq 0,\ i = 1, \ldots, n_d$ 且 $\sum_{i=1}^{n_d} \theta_i = 1$。借助以下三个命题，我们研究公式（14）中概率分布可行集的几何结构。

命题 1 *Let* $(X_1, \ldots, X_d)$ *be a multivariate Bernoulli random variable with distribution* $p = p_\alpha,\ \alpha \in \mathcal{D}$ *uniform margins and given second-order moments* $\mu_{12}\ \mu_{13}\ \ldots\ \mu_{d-1\,d}$. *Then the multivariate Bernoulli random variable* $X'_1\ \ldots\ X'_d$ *with distribution* $p' = p'_\alpha = p_{1-\alpha},\ \alpha \in \mathcal{D}$ *also has uniform margins and the same second-order moments.*

Proof 设 $(X'_1, \ldots, X'_d)$ 为一个分布为 p' 的多元伯努利随机变量。我们得到

$$\mu'_i = E[X'_i] = \sum_{\alpha \in \mathcal{D}, \alpha_i = 1} p'_\alpha = \sum_{\alpha \in \mathcal{D}, \alpha_i = 1} p_{1-\alpha} = \sum_{\alpha \in \mathcal{D}, \alpha_i = 0} p_\alpha = 1 - E[X_i] = \frac{1}{2},\ 1, \ldots, d.$$

由此可知，$(X'_1, \ldots, X'_d)$ 具有均匀边际。类似地，对于二阶矩，我们得到

$$\mu'_{ij} = E[X'_i X'_j] = \sum_{\alpha \in \mathcal{D}, \alpha_i = \alpha_j = 1} p'_\alpha = \sum_{\alpha \in \mathcal{D}, \alpha_i = \alpha_j = 1} p_{1-\alpha} = \sum_{\alpha \in \mathcal{D}, \alpha_i = \alpha_j = 0} p_\alpha =$$
$$= E[(1 - X_i)(1 - X_j)] = 1 - E[X_i] - E[X_j] + E[X_i X_j] = E[X_i X_j] = \mu_{ij},$$
$$i, j = 1, \ldots, d, i < j.$$

□

在下面的结果中，我们证明可行解的对称性可以扩展到极端概率质量函数，因此极端概率质量函数成对出现，并具有精确的组合结构。证明基于极射线的几何定义，见 [20,第 65] 页。

命题 2 *If* $r = (r_\alpha)$ *is an extreme pmf of the polyhedral cone for the problem inEq.* (13) *then also* $r' = (r_{1-\alpha})$ *is an extreme pmf.*

Proof 为简便起见，我们研究由式（13）定义的未归一化多面锥。根据命题 1，我们知道 r' 是一个可行分布。我们需要证明 r' 是一个极端概率质量函数。我们知道，当且仅当写成 r 是 $\mathcal{F}(H_d)$ 中的一个极端概率质量函数

$$r = p_1 + p_2 \qquad p_1, p_2 \in \mathcal{C}(H_d)$$

那么 $p_1 = \theta_1 r$ 和 $p_2 = \theta_2 r, \theta_1,\ \theta_2 \geq 0$。

现在，考虑 r' 并用反证法进行证明。假设 r' 不是一个极端概率质量函数。因此，我们可以写成

$$r' = p_1 + p_2 \qquad p_1, p_2 \in \mathcal{C}(H_d) \tag{15}$$

其中 p_1 和 p_2 并非同时是 r' 的倍数。再次对前面的等式进行对称化，并注意到 $(r')' = r$，我们得到

$$r = (r')' = p'_1 + p'_2$$

由于 p'_1 和 p'_2 属于 $\mathcal{C}(H_d)$，根据命题 1，且 r 是一个极端概率质量函数，因此我们得到 p'_1和 p'_2 都必须是 r的非负倍数：

$$p'_1 = \theta_1 r, \qquad p'_2 = \theta_2 r$$

对于某个非负的 θ_1，θ_2。再次对指标进行对称化，我们得到：

$$p_1 = \theta_1 r', \qquad p_2 = \theta_2 r'$$

因此二者 p_1 和 p_2 都是 r'的非负倍数，这与公式（15）矛盾 . □

最后一个结果涉及可行集中的概率质量函数在对数线性表示下的对称性。这种对称性源自该模型零均值参数化的特殊性质。回顾一下，将一个非负概率分布写成饱和对数线性表示时，我们有

$$\log p_{i_1 \dots i_d} = \lambda + \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} \lambda^S_{i_S},$$

其中求和遍历所有非空子集 S，这些子集属于 $\{1, \dots, d\}$，而每个 i_S都是由 S所标识的指标子集。零均值参数化假设参数 $\lambda^S_{i_S}$ 在 S中的所有一维指标上求和为零。例如，在双向情形下，我们写作：

$$\log p_{\alpha_1 \alpha_2} = \lambda + \lambda^1_{\alpha_1} + \lambda^2_{\alpha_2} + \lambda^{12}_{\alpha_1 \alpha_2}$$

其中 $\lambda^i_0 + \lambda^i_1 = 0$、 $i = 1$，2 和 $\lambda^{12}_{00} + \lambda^{12}_{01} = \lambda^{12}_{10} + \lambda^{12}_{11} = \lambda^{12}_{00} + \lambda^{12}_{00} = \lambda^{12}_{01} + \lambda^{12}_{11} = 0$。

命题 3 *Using the zero-mean parametrization of the saturated log-linear model, two nonnegative symmetric pmfs p and p' have the same log-linear parameters in absolute value with the same sign for even orders of interaction and with the opposite sign for odd orders of interaction.*

Proof 只需考虑 p 的饱和对数线性形式：

$$\log p_{i_1 \dots i_d} = \lambda + \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} \lambda^S_{i_S},$$

注意到将 p 映射到 p' 会改变 α 中的所有指标，因此可以立即得出

$$\log p_{i_1 \dots i_d} = \lambda + \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|} \lambda^S_{i_S},$$

由此得证。 □

我们指出，上述命题要求分布严格为正。然而，可以通过加上一个较小的正数来避免零概率，从而将其应用于极端概率质量函数。我们将在下面的例子中看到这个问题。

Example 1 考虑三维列联表 T_0 其概率质量函数 p_0 见表 1 以及其边际优势比 ω^{12}_M， ω^{13}_M和 ω^{23}_M。

我们的目标是确定具有均匀边际且边际优势比相同的概率质量函数所构成的凸多胞体。如前文所述，对于二维列联表，存在

边际优势比 ω_M^{12} 与具有均匀边际的概率质量函数中对应于 (1，1) 单元格的取值 p_{11} 之间存在一一对应关系：

$$p_{11} = \frac{\sqrt{\omega_M^{12}}}{2\left(\sqrt{\omega_M^{12}}+1\right)}. \tag{16}$$

对于三维列联表，给定 ω_M^{12}、ω_M^{13}和 ω_M^{23}，方程（16）可用于分别确定二阶矩 $\mu_{12} = p_{110}+p_{111}$、$\mu_{13} = p_{101}+p_{111}$和 $\mu_{23} = p_{011}+p_{111}$。在我们的例子中，得到：

$$\mu_{12} = 0.194, \qquad \mu_{13} = 0.222, \qquad \mu_{23} = 0.257.$$

构建相应的 H_3 矩阵并使用 `4ti2`, [21], ，我们得到极射线。经归一化后，这些极射线给出该多面体的极端概率质量函数；该多面体包含所有具有均匀边际且边际优势比与原始表相同的概率质量函数 p_0。极端概率质量函数列于表 2。在三维二元情形中，可行集是单纯形中的一条线段，其极端概率质量函数就是线段的两个端点。

将可行集中的一个一般概率定义为混合分布 $p = \alpha r_1 + (1-\alpha) r_2$，其中 $\alpha \in [0,1]$。容易验证，三维优势比

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

当 α 趋近于 1 时趋近于 0，而当 α 趋近于 0 时则发散至 $+\infty$，从而涵盖所有可能的三维依赖结构。

此外，表中两个极端概率质量函数2 按照命题2. 所述的意义具有对称性。为了阐明命题 3中所述对数线性表示的对称性，我们考察了以下对数线性表示

$$\log r_{\alpha_1\alpha_2\alpha_3} = \lambda + \lambda^1_{\alpha_1} + \lambda^2_{\alpha_2} + \lambda^3_{\alpha_3} + \lambda^{12}_{\alpha_1\alpha_2} + \lambda^{13}_{\alpha_1\alpha_3} + \lambda^{23}_{\alpha_2\alpha_3} + \lambda^{123}_{\alpha_1\alpha_2\alpha_3},$$

其中，$r_{i,\alpha}$ 的值设为 $r_{i,\alpha} + 10^{-8}$，以避免计算对数时出现数值问题。除了零均值参数化之外，我们还考察了角参数化，以说明这种对称性是零均值参数化特有的现象。角参数化是一种成熟的方法，可用于获得满秩模型矩阵。它假设除 $i_S = 1$ 外的所有参数均为零（即 S 中的所有指标都等于 1）。在这个三维例子中，这意味着只有 λ、λ^1_1、λ^2_1、λ^3_1、λ^{12}_{11} λ^{13}_{11} λ^{23}_{11} λ^{13}_{11} λ^{23}_{11} λ^{123}_{111} 、、、和 均非零。

两种参数化方式中，与 r_1 和 r_2 相关的对数线性参数值报告于表3。

如果固定观测边际而非均匀边际，对称性就会消失。现在，我们的目标是确定具有相同边际和边际优势比的概率质量函数的凸多胞体： p_0。构造相应的 H_3 矩阵，并再次使用`4ti2`, [21], 我们得到极射线；经过归一化后，这些极射线给出了相关可行集的极端概率质量函数。极端概率质量函数列于表4。

表1 表 T_0 示例1 报告 p_0 （精确值）及其边际优势比（四舍五入至小数点后两位）。数据来自 [16]。

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

表 2 极端概率质量函数 r_1 和 r_2 具有均匀边际和边际优势比的概率质量函数凸多胞体 p_0 示例 1，所有数值均四舍五入至小数点后三位。

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

表 3 极端概率质量函数的对数线性表示的参数 r_1 和 r_2，见表 2，所有数值均四舍五入至小数点后两位。

参数化	pmf	λ	λ_1^1	λ_1^2	λ_1^3	λ_{11}^{11}	λ_{11}^{13}	λ_{11}^{23}	λ_{111}^{111}
角点	r_1	−18.42	17.06	16.92	16.78	−19.41	−18.42	−17.75	21.49
角点	r_2	−1.76	−0.72	−1.24	−2.10	2.08	3.07	3.74	−21.49
零均值	r_1	−4.25	1.76	1.85	2.03	−2.17	−1.92	−1.75	2.69
零均值	r_2	−4.25	−1.76	−1.85	−2.03	−2.17	−1.92	−1.75	−2.69

表 4 极端概率质量函数 r_1 和 r_2 具有相同边际及边际优势比的概率质量函数凸多胞体的 p_0 示例1，所有数值均四舍五入至小数点后三位。

pmf	000	001	010	011	100	101	110	111
r_1	0.050	0.100	0.350	0.150	0.150	0	0.100	0.100
r_2	0.150	0	0.250	0.250	0.050	0.100	0.200	0

即使利用本节命题中确立的对称性，在可行集中选取特定分布仍颇具挑战。当维数 d 适中，即不超过 6 或 7 时，所提出的方法在计算上可行：可以列出极端概率质量函数，并使用基于三角剖分的方法对相应多面体进行均匀抽样，例如 R 软件包 `uniformly` [22] 中提供的方法。随着维数 d增加，极端概率质量函数的数量变得过多时，可以使用其他技术，例如 R 软件包 `volesti` [23]，中实现的技术；该工具不依赖三角剖分，而是采用基于随机游走的方法，从给定多面体中进行均匀抽样。 [24] 中给出了低维和高维应用示例。在高维情形下，另一种可行方法是使用 MCMC 类方法。例如，代数统计学中的马尔可夫基已成功用于相关问题，以便在概率单纯形 [25] 的子集中进行遍历。不过，对这些技术的全面研究超出了本文的范围。

表5 第4.1节示例的四维列联表。数据改编自 [26]。

水的软硬度	品牌偏好	M 的先前使用者		M 的先前非使用者	
		高T	低T	高T	低T
Soft	X	19	57	29	63
	M	29	49	27	53
中等+困难	X	47	84	75	134
	M	90	107	53	92

4 示例与应用

本节将说明所提出的框架如何应用于真实和合成的列联表。我们首先分析一个四维二元列联表，展示如何通过其极射线刻画具有固定两两依赖和均匀边际的分布可行集。随后，我们探讨观测零值的存在如何影响可行区域的几何形状，并可能限制或改变依赖结构。最后，我们介绍一个评分者一致性方面的真实数据应用，说明极端解如何帮助解释高阶依赖并指导模型选择。

4.1 四维二元列联表：几何结构与极射线

在此前说明了 $d = 3$这一情形后，我们现在考察一个四维二元列联表，以展示所提出的方法如何扩展到高维。这个示例改编自 [26], ，使我们能够显式构造矩阵 H_4，确定相应的多面锥并计算其极射线。与第3节讨论的三维情形不同，四维情形呈现出更复杂的几何结构，极射线数量显著增加，高阶依赖模式也更加复杂。

[26] 中给出的原始表的维数为 3×2^3。我们通过合并第一个变量的两个水平，将其调整为适用于我们框架的形式。表5列出了所得的表，总结了对1，008名消费者按四个变量进行的交叉分类：所用洗衣用水的软硬程度（软水、中等+硬水）；此前是否使用过洗涤剂 M；所用洗衣用水的温度（高、低）；对洗涤剂 X 相较于 M 的偏好。

概率质量函数 p_0 对应于表5 的四维列联表通过归一化计算得到。

与 p_0 对应的边际优势比列于表 6 中。与我们处理三维列联表时的做法类似，给定 ω_M^{ij}，可利用方程 Eq. (16) 确定相应的二阶矩 μ_{ij}, $i, j \in \{1, 2, 3, 4\}, i < j$，从而得到满足均匀边际且边际优势比与 p_0 相同的概率质量函数所需的条件。这些二阶矩列于表 6 中。构建相应的 H_4 矩阵并使用 `4ti2`, [21], ，我们得到极射线；经过

归一化后，这些极射线给出了多面体的极端概率质量函数，该多面体包含所有具有均匀边际且边际优势比与原始表 p_0 相同的概率质量函数。我们得到 $n_d = 96$ 个极端概率质量函数。

表 6 的边际优势比 p_0 以及具有均匀边际且边际优势比相同的概率质量函数的二阶矩p_0 （表5），所有数值均四舍五入至小数点后三位。

i	j	ω_M^{ij}	μ_{ij}
1	2	1.070	0.254
2	3	0.563	0.214
1	3	0.966	0.248
3	4	1.158	0.259
2	4	0.761	0.233
1	4	0.737	0.231

通过对这些极端概率质量函数进行凸组合，我们可以得到丰富多样的可行联合分布，它们都具有相同的成对依赖结构，从而能够更灵活地选择最适合进一步分析的模型。

4.2 评分者一致性问题的实证应用

下面，我们将通过涉及三名评分者的真实一致性数据，说明所提出的框架如何应用于实践。在这类情境中，人们通常不仅关注成对一致性，还希望识别仅凭成对测度无法捕捉的高阶一致性模式。表7所示数据被视为一个二元表，位于 [27], ，该表由原文 [28]中报告的三分类表合并类别后得到。在这项关于说服性沟通的研究中，三名评分者将 164 名患者的认知反应分为两类：“正面或中性”（第 2 级）和“负面”（第 1 级）。

如前所述，标准方法（如 MIPFP 或对数线性模型）会返回单个列联表，并且往往会隐含地令高阶交互作用为零；当实际存在高阶依赖时，这可能导致误导性的解释。例如，表 7 中的三阶交互作用可用三维优势比来描述，其观测值为

$$\widehat{\omega}^{123} = \frac{n_{111}n_{122}n_{212}n_{221}}{n_{222}n_{211}n_{121}n_{112}} = 2.96625.$$

表7 评分者一致性问题中的表格。数据来自 [27]。

		R3	
R1	R2	1	2
1	1	113	5
	2	5	7
2	1	4	3
	2	3	24

表 8 p_0 及其边际优势比，表7中的所有数值均四舍五入至小数点后三位。

pmf	000	001	010	011	100	101	110	111	ω_M^{12}	ω_M^{13}	ω_M^{23}
p_0	0.689	0.03	0.03	0.043	0.024	0.018	0.018	0.146	37.929	37.929	56.672

表 9 极端概率质量函数 r_1 和 r_2 对于 p_0 表8中的所有数值均四舍五入至小数点后三位。

pmf	000	001	010	011	100	101	110	111
r_1	0.372	0.059	0.059	0.011	0.07	0	0	0.43
r_2	0.43	0	0	0.07	0.011	0.059	0.059	0.372

这一数值表明，除二元相关性外还存在显著的依赖模式，因此，在这类框架中，使用例如基于 MIPFP 的算法（该算法将高阶优势比设为 1）进行模拟并不合理。相反，通过刻画所有具有均匀边际和固定两两依赖的可行概率质量函数，我们的框架能够在保持成对一致性不变的同时，考察高阶交互作用如何变化。这为一致性结构提供了更丰富、更透明的视角，并有助于判断观测到的依赖关系能否仅由成对关联解释，还是存在三方一致性效应。与表 p_0 中三维列联表相对应的概率质量函数见表7 ，见表 8。

按照前几节讨论的步骤，我们可以推导出极端概率质量函数，列于表9。两种参数化方式中与 r_1 和 r_2 相关的参数值列于表10。为避免计算对数时出现数值问题，$r_{i,\alpha}$的值设为$r_{i,\alpha} + 10^{-8}$ 。

表 10 极端概率质量函数的参数 r_1 和 r_2 在表 9中，所有数值均四舍五入至小数点后两位。

参数化	pmf	λ	λ_1^1	λ_1^2	λ_1^3	λ_{11}^{11}	λ_{11}^{13}	λ_{11}^{23}	λ_{111}^{111}
角点	r_1	-0.99	-1.67	-1.85	-1.85	-13.91	-13.91	0.19	33.14
角点	r_2	-0.84	-3.65	-17.58	-17.58	19.23	19.23	33.34	-33.14
零均值	r_1	-6.44	-3.65	-0.21	-0.21	0.66	0.66	4.19	4.14
零均值	r_2	-6.44	3.65	0.21	0.21	0.66	0.66	4.19	-4.14

5 讨论与未来工作

本文提出了一个几何框架，用于刻画具有均匀边际且固定成对依赖关系（以相关系数或边际优势比表示）的 d 个二元变量的全部联合分布。不同于现有方法（如 MIPFP 或基于对数线性模型的抽样）通常选择单个代表性分布，本文描述了所有可容许解构成的空间。我们证明该可行集构成一个多面锥，研究其极射线，并证明了若干对称性。通过识别这些射线，我们得以深入了解在保持成对信息固定的同时，高阶交互作用如何变化。这一视角有助于在生成列联表时，更审慎、灵活地选择依赖结构。

现在，我们讨论本文所提方法的一些局限性。假设边际均匀虽然在数学和解释层面都具有优势，但这一选择也可能带来限制，尤其是在数据中存在结构零或抽样零时。零项可能大幅缩小分布的可行集；在某些情况下，甚至无法在保留原有支撑集的同时，构造具有均匀边际和指定边际优势比的列联表。在这种情况下，理解零模式与均匀边际之间的相互作用至关重要。 [7, 16], 对此方面进行了研究，并给出了兼容性条件。这些结果是我们框架的基础，可对其进行调整，以确定可行集何时非空，并解释零模式对可容许分布结构的影响。

从计算角度来看，仍有若干挑战尚待解决。正如第3节所述，尽管已有从凸多胞体中抽取采样点的成熟方法，但这些方法是否适用于我们特定的几何结构，以及其效率如何，尚未得到充分探索。评估并调整这些抽样策略以适应我们的框架，是未来研究的一个有前景的方向，我们将在后续研究中探讨这一问题。此外，本文构建的框架聚焦于二元表。未来研究的一个自然方向是将该框架扩展至多级分类变量，这类变量在实际应用中十分常见。此类多级设置会引入规模更大、结构更复杂的可行区域，从而使几何刻画更具挑战性，也更令人感兴趣。我们相信，这将进一步拓宽我们方法的适用范围。

声明

作者声明不存在利益冲突。

参考文献

[1]Agresti, A.：《分类数据分析》。第三版。约翰·威立父子出版公司，霍博肯，新泽西州（2012）[2] Rudas, T.：《分类数据分析讲义》。施普林格，纽约（2018）[3] Eye, A., Mun, E.Y.：《评定者一致性分析》。显变量方法。劳特利奇，纽约（2012）[4] Fontana, R., Semeraro, P.：给定一组指定矩的多元伯努利分布表示。多元分析杂志，168，290–303（2018）https://doi.org/10.1016/j.jmva.2018.08.003[5] Hammond, C., Heijden, P.G.M., Smith, P.A.：生成边缘概率固定且依赖结构由对数线性模型描述的列联表。统计计算与模拟杂志，94(12)，2797–2812（2024）https://doi.org/10. 1080/00949655.2024.2353760[6] Barth´elemy, J., Suesse, T.：mipfp：用于多维数组拟合及多元伯努利分布模拟的 R 软件包。统计软件杂志，86（2018）https://doi.org/10.18637/jss.v086.c02[7] Fontana, R., Perrone, E., Rapallo, F.：零模式对三维二元表依赖性的影响。载于：Boccuzzo, G., Bovo, E., Manisera, M., Salmaso, L.（编）《IES 2025——创新与社会：用于评估和质量的统计学与数据科学——短篇论文集》，第826–832页。CLEUP，帕多瓦，意大利（2025）[8] Perrone, E., Solus, L., Uhler, C.：离散联结函数的几何。多元分析杂志，172，162–179（2019）https://doi.org/10.1016/j.jmva.2019.01.014[9] Geenens, G.：离散随机向量的联结函数建模。依赖模型，8(1)，417–440（2020）https://doi.org/10.1515/demo-2020-0022[10] Perrone,E., Durante, F.：离散联结函数多胞体的极点。载于：第19届国际模糊系统协会世界大会（IFSA）、欧洲模糊逻辑与技术学会第12届会议（EUSFLAT）及第11届聚合算子国际暑期学校（AGOP）联合论文集，第596–601页（2021）。https://doi.org/10.2991/asum.k.210827.080[11] Perrone, E.：离散联结函数多胞体及其应用。载于：Garc´ıa-Escudero, L.A., Gordaliza, A., Mayo, A., Lubiano Gomez, M.A., Gil, M.A., Grzegorzewski,

P., Hryniewicz, O.（编）《搭建数据科学中软方法与统计方法之间的桥梁》，第319–325页。施普林格，查姆（2023）[12] Perrone, E., Fontana, R., Rapallo, F.：具有均匀边际的多维列联表。载于：Ansari, J., Fuchs, S., Trutschnig, W., Lubiano, M.A., Gil, M.A., Grzegorzewski, P., Hryniewicz, O.（编）《结合、建模与分析不精确性、随机性和依赖性》，第349–356页。施普林格，查姆（2024）[13] Kojadinovic, I., Martini, T.：具有矩形支撑集的离散二元分布的类 copula 推断。电子统计学杂志，18(1)，2571–2619（2024）https://doi.org/10.1214/24-EJS2261 [14] Feinstein, A.R., Cicchetti, D.V.：高一致性但低卡帕系数：I. 两个悖论的问题。临床流行病学杂志，43(6)，543–549（1990）https://doi.org/10.1016/0895-4356(90)90158-L[15] Agresti, A., Ghosh, A., Bini, M.：迭代比例拟合卡帕系数：描述边际分布对一致性度量的潜在影响。生物统计学杂志，7，811–820（1995）[16] Fontana, R., Perrone, E., Rapallo, F.：具有均匀边际的多维二元列联表中的零模式。国际近似推理杂志，188，109582（2026）https://doi.org/10.1016/j.ijar.2025.109582[17] Lee, A.J.：生成具有固定边际分布和指定关联程度的随机二元偏差变量。美国统计学家，47(3)，209–215（1993）[18] Fienberg, S.E.：列联表估计的迭代程序。数学统计年刊，41(3)，907–917（1970）https://doi.org/10.1214/aoms/1177696968[19] Dahl, G.：《凸性、多面体理论与组合优化导论》。技术报告 Kompendium 67 IN 330，信息学系，奥斯陆大学（1997）[20] Barvinok, A.：《凸性教程》。数学研究生学习丛书，第54卷。美国数学会，普罗维登斯，罗得岛州（2002）[21] 4ti2 团队：4ti2——用于线性空间上代数、几何和组合问题的软件包。https://4ti2.github.io[22] Laurent, S.：Uniformly：均匀抽样。（2018）。R 软件包版本 0.1.0。https: //CRAN.R-project.org/package=uniformly[23] Fisikopoulos, V., Chalkis, A.：Volesti：凸多面体的体积近似与抽样。（2020）。R 软件包版本 1.1.2。https://CRAN.R-project.org/ package=volesti [24] Fontana, R., Semeraro, P.：可交换伯努利分布：高维模拟、估计与检验。《统计规划与推断杂志》

[25] Pistone, G., Rapallo, F., Rogantin, M.P.：有限空间中的康托罗维奇问题与列联表变动的马尔可夫链蒙特卡罗方法。电子统计学杂志 **15**(1), 880–907 (2021) https://doi.org/10.1214/21-EJS1804[26]Goodman, L.A.：多维列联表分析：用于构建多个分类模型的逐步程序和直接估计方法。Technometrics **13**(1), 33–61 (1971) [27] Vanbelle, S.：二元量表上多个评定者之间一致性的统计推断。Br. J. Math. Stat. Psychol. **77**, 245–260 (2023) https://doi.org/10:1111/bmsp.12333[28] Mart´ın Andr´es, A., ´Alvarez Hern´andez, M.：重新审视 Hubert 多评定者卡帕系数。Br. J.Math. Stat.Psychol. **73**, 1–22 (2020) https://doi.org/10:1111/bmsp.12167