arXiv:2603.29786v1 [stat.OT] 31 Mar 2026

证据支持的对称性

Grant Molnar

2026年4月1日

摘要

对于事件 A 和 B，我们有

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \qquad \Longleftrightarrow \qquad \mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$$

只要这四个量都有定义，结论便成立。换言之，B 是 A 的证据，当且仅当 A 是 B 的证据。本文通过交叉相乘、协方差、耦合参数、优势比、点互信息、组合双重计数和混合离散导数，给出了这一事实的七种不同证明，并围绕它展开了一系列解读。一旦边缘分布 $\mathbb{P}(A)$ 和 $\mathbb{P}(B)$ 固定，一个 2×2 表格就只有一个自由度，因此，所有标量的正关联度量都必须由同一个带符号参数决定。

目录

1 引言 **2**

2 代数证明 **4**

2.1 视角.4 2.2 证明.4 2.3 行列式与秩一.5 2.4 有向面积与几何.6

3 协变 性证明 **6**

3.1 视角.6 3.2 证明.6 3.3 回归斜率.7 3.4 沃尔什–傅里叶展开.7 3.5 希尔伯特空间与典型相关视角.7

4 耦合 g 参数证明 **8**

4.1 视角.8 4.2 证明.8 4.3 归一化交互参数.9 4.4 微分观点.10

4.5 一致性与不一致性.10 4.6 传输与不匹配成本.11

5 优势比证明 **11**
5.1 视角.11 5.2 证明.11 5.3 证据的贝叶斯互易性.11 5.4 伊辛相互作用与指数族.12 5.5 单调似然比.12 5.6 逻辑回归.12

6 信息论证明 13
6.1 视角.13 6.2 证明.13 6.3 惊异度降低.14 6.4 互信息与依赖方向.14

7 组合证明 **14**
7.1 视角.14 7.2 证明.15 7.3 推广到任意概率空间.15 7.4 随机对决解释.15

8 对数势 性证明 16
8.1 视角.16 8.2 证明.16 8.3 全正性与 FKG不等式.17 8.4 随机序.17 8.5 对数线性模型.17

9 结论 **17**

1 引言

Thro在本文中， A和 B 是环境概率空间中的事件，且 h

$$0 < \mathbb{P}(A) < 1 \qquad \text{and} \qquad 0 < \mathbb{P}(B) < 1,$$

因此，四个条件概率 $\mathbb{P}(A \mid B)$、 $\mathbb{P}(A \mid \neg B)$、 $\mathbb{P}(B \mid A)$、 $\mathbb{P}(B \mid \neg A)$ 都有定义。

将四个联合概率写作

$$p := \mathbb{P}(A \wedge B), \quad q := \mathbb{P}(A \wedge \neg B), \quad r := \mathbb{P}(\neg A \wedge B), \quad s := \mathbb{P}(\neg A \wedge \neg B),$$

因此 $p+q+r+s = 1$。边缘分布为 $\mathbb{P}(A) = p+q$和 $\mathbb{P}(B) = p+r$。定义 $\Delta := ps - qr$。

这篇笔记的核心观点是，二元关联是一维的。一旦固定了 $\mathbb{P}(A)$ 和 $\mathbb{P}(B)$，一个 2×2 表格只有一个自由参数，因此，所有衡量正关联与负关联的标量指标，都必须在同一个阈值处改变符号。

定理 1.1（证据支持的对称性）。 以下命题等价。

(i) $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$。 ***(ii)*** $\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$。 ***(iii)*** $\Delta > 0$。

将每个“$>$”替换为“$=$”或“$<$”，即可得到另外两组等价关系，因此 Δ 的三种可能符号刻画了二元关联的三种可能定性状态。

定理 1.1 是经典结果，这里给出的若干证明也是标准证明或民间流传的证明；本文的贡献在于将它们汇集起来，并围绕这样一个观察进行组织：一旦边缘分布固定，二元关联就是一维的。

注 1.2（两种二元编码）。 我们将使用事件 A 和 B 的两种标准数值编码。默认采用指示变量编码

$$X := \mathbf{1}_A, \qquad Y := \mathbf{1}_B,$$

因此 $X,\ Y \in \{0, 1\}$。对于条件概率、指示变量的协方差和回归公式，这是自然的选择。在沃尔什–傅里叶和伊辛模型的视角下，使用自旋编码则更为方便

$$U := 2X - 1, \qquad V := 2Y - 1,$$

因此 $U,\ V \in \{-1, +1\}$。当我们从一种编码转换到另一种编码时，总会明确标示。

在证明定理 1.1 之前，我们先举几个例子说明它的重要性。

例 1.3（医学检测）。 设 A 表示患者患有某种疾病这一事件， B 表示筛查测试结果呈阳性这一事件。如果该测试具有信息价值——也就是说， $\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$，即真阳性率高于假阳性率——那么定理 1.1 保证，阳性结果会使患病的概率高于阴性结果。

例 1.4（相关与预测）。 设 A表示一名学生在数学考试中的成绩高于中位数这一事件， B 表示同一名学生在物理考试中的成绩高于中位数这一事件。如果 A 使 B 更有可能，那么 B 也会使 A 更有可能。“数学预测物理”和“物理预测数学”这两种预测要么同时成立，要么同时不成立。

例 1.5（配对性能测试）。 设 A 表示某制造部件通过热应力测试这一事件， B 表示该部件通过振动测试这一事件。如果通过热应力测试会使通过振动测试的可能性增大，那么通过振动测试也同样会使通过热应力测试的可能性增大。这种关联在两个预测方向上都成立。

本文余下部分从不同视角给出七个证明。第2节从直接的代数化简入手，导出标量不变量 $\Delta = ps - qr$。第3节用协方差来重新表述符号条件，并进一步以更具概念性的希尔伯特空间语言加以阐释。第4节则给出整体结构上的解释：边缘分布一旦固定，就只有一个交互坐标。第5节通过优势比和贝叶斯因子从统计学角度解释这一现象；第6节则用点互信息和惊异度降低来表述其信息论形式。最后，第7节从组合角度重新表述同一符号条件，第8节则借助离散势理论加以阐释。第9节梳理前述结果，并强调为何在 2×2情形下，许多标量关联概念都会归结为相同的符号。

2 代数证明

2.1 视角

本节记录我们对定理 1.1 最直接的证明。两个条件概率比较经初等代数化简后，都归结为单个标量不等式 $ps - qr > 0$，因此类行列式量 $\Delta = ps - qr$ 立即显现为该2×2 表格的基本不变量。

2.2 证明

定理 *1.1* 的代数版本证明。我们证明 (i) 和 (ii) 都等价于 (iii)。

对于 (i) $\Leftrightarrow$ (iii)，记

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff \frac{p}{p+r} > \frac{q}{q+s}.$$

两个分母 均为正数，因此交叉相乘后不等号方向不变，得到

$$p(q+s) > q(p+r) \iff ps > qr \iff \Delta > 0.$$

对于 (ii) $\Leftrightarrow$ (iii)，令

$$\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A) \iff \frac{p}{p+q} > \frac{r}{r+s}.$$

交叉相乘可得

$$p(r+s) > r(p+q) \iff ps > qr \iff \Delta > 0.$$

我们的论断由此得出。 □

2.3 行列式与秩为 1

将联合概率排列在矩阵中

$$M \coloneqq \begin{pmatrix} p & q \\ r & s \end{pmatrix}.$$

于是 $\det M = ps - qr = \Delta$。现在可以将代数证明重新解读为关于 $\det M$ 的陈述。

命题 2.1。 A 与 B 独立，当且仅当 M 的秩为 1。

证明。 如果 A 与 B 独立，那么 M 可以分解为边缘向量的外积，

$$M = \begin{pmatrix} \mathbb{P}(A) \\ \mathbb{P}(\neg A) \end{pmatrix} \begin{pmatrix} \mathbb{P}(B) & \mathbb{P}(\neg B) \end{pmatrix}.$$

Any 外积的秩至多为 1，并且由于 $M \neq 0$，秩恰好为 1。

反过来，如果 $\operatorname{rank}(M) = 1$，那么第二行是第一行的标量倍数：$(r,\ s) = \lambda\,(p,\ q)$，其中某个 $\lambda \geq 0$。因此

$$\mathbb{P}(B) = p + r = (1+\lambda)p, \qquad \mathbb{P}(A) = p + q.$$

Also

$$1 = p + q + r + s = (1+\lambda)(p+q),$$

因此 $(1+\lambda)(p+q) = 1$。因此

$$\mathbb{P}(A)\mathbb{P}(B) = (p+q)(1+\lambda)p = p = \mathbb{P}(A \wedge B).$$

因此，A 和 B 相互独立。由于 M 是一个元素非负且总和为 1 的 2×2，其秩为 1 或 2，因此秩为 1 等价于 $\det M = 0$。 □

考虑条件概率差

$$\mathbb{P}(A \mid B) - \mathbb{P}(A \mid \neg B) = \frac{\det M}{(p+r)(q+s)}, \tag{1}$$

$$[4pt]\mathbb{P}(B \mid A) - \mathbb{P}(B \mid \neg A) = \frac{\det M}{(p+q)(r+s)}, \tag{2}$$

它们的分母严格为正，且分子相同。这是因为 (2) 是将 (1) 应用于 $M^{\top}$ 后得到的，并且 $\det M^{\top} = \det M$。

注释 2.2（代数统计学）。在代数统计学的语言中 [5]，2×2 列联表的独立模型是簇 $ps - qr = 0$；这是塞格雷簇最简单的实例。对于 $m \times n$ 表格，独立性等价于所有2×2 子式都为零。二元情形很特殊，因为只有一个子式。

2.4 有向面积与几何

M 的两行是 $\mathbb{R}^2$中的向量,

$$\mathbf{r}_A \coloneqq (p, q), \qquad \mathbf{r}_{\neg A} \coloneqq (r, s),$$

并且 Δ 是它们张成的平行四边形的有向面积。当 $\Delta = 0$ 各行成比例时——给定 A时 B 的条件分布与给定 $\neg A$时相同——这意味着独立。当 $\Delta > 0$时，向量$\mathbf{r}_A$ 相对于$\mathbf{r}_{\neg A}$"向第一个坐标倾斜"，也就是说， A 比 $\neg A$ 更集中于 B 。对 M 转置不会改变行列式，因此列的比较也遵循同样的倾斜关系。在外代数中， $\Delta = \mathbf{r}_A \wedge \mathbf{r}_{\neg A}$，定理 1.1由此得出，因为 $\det M^\top = \det M$。

3 协方差证明

3.1 视角

本节用希尔伯特空间的语言重新表述定理 1.1。关键在于，两个条件方向都由单个双线性量$\mathrm{Cov}(\mathbf{1}_A, \mathbf{1}_B)$的符号决定，或者等价地说，由与之相关的一维相关参数决定。

3.2 证明

令 $X = \mathbf{1}_A$ 和 $Y = \mathbf{1}_B$ 与通常一样。

命题 3.1. 我们有等式 $\mathrm{协方差}(X, Y) = \Delta$。

证明。 我们计算

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = p - (p+q)(p+r) = p(1-p-q-r) - qr = ps - qr.$$

□

命题 3.2. 对于任意满足 $0 < \mathbb{P}(B) < 1$ 的事件 A 和 B，

$$\mathbb{P}(A \mid B) - \mathbb{P}(A) = \frac{\mathrm{Cov}(X, Y)}{\mathbb{P}(B)}.$$

证明。 我们计算

$$\mathbb{P}(A \mid B) - \mathbb{P}(A) = \frac{\mathbb{P}(A \wedge B)}{\mathbb{P}(B)} - \mathbb{P}(A) = \frac{\mathbb{P}(A \wedge B) - \mathbb{P}(A)\,\mathbb{P}(B)}{\mathbb{P}(B)} = \frac{\mathrm{Cov}(X, Y)}{\mathbb{P}(B)}.$$

□

定理 *1.1* 的证明，协方差版本。 根据命题 3.2， $\mathbb{P}(A \mid B) > \mathbb{P}(A)$ 当且仅当 $\mathrm{Cov}(X, Y) > 0$。将 $\neg B$ 替换为 B 应用同一恒等式，并注意到 $\mathrm{Cov}(X, \mathbf{1}_{\neg B}) = -\mathrm{Cov}(X, Y)$，可得 $\mathbb{P}(A \mid \neg B) < \mathbb{P}(A)$ 当且仅当 $\mathrm{Cov}(X, Y) > 0$。结合这些恒等式，我们得出

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff \mathrm{Cov}(X, Y) > 0.$$

由于 $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$，将 A 与 B 交换后，同样的等价关系仍然成立 . □

由此得到一个有用的恒等式

$$\mathbb{P}(A \mid B) - \mathbb{P}(A \mid \neg B) = \operatorname{Cov}(X, Y) \left(\frac{1}{\mathbb{P}(B)} + \frac{1}{\mathbb{P}(\neg B)} \right).$$

3.3 回归斜率

X对 Y进行最小二乘回归的斜率为 $\beta_{X|Y} \coloneqq \mathsf{Cov}(X,\ Y)\ /\mathsf{Var}(Y)$，其符号与 $\mathsf{Cov}(X,\ Y)\ = \Delta$ 的符号相同。Y对 X进行回归的斜率为 $\beta_{Y|X} \coloneqq \mathsf{Cov}(X,\ Y)\ /\mathsf{Var}(X)$，符号也相同。因此，“ $\mathbf{1}_A$ 的条件均值随 $\mathbf{1}_B$增加而增加”与反过来的说法是同一断言。

3.4 Walsh–Fourier 展开

将事件编码为变量 ± 1 和 $U = 2X - 1$ 。令

$$f(u, v) \coloneqq \mathbb{P}(U = u, V = v)$$

为联合　　概率质量函数，定义在 $\{-1, +1\}^2$上。相对于沃尔什基

$$1, \quad u, \quad v, \quad uv$$

for $\{-1, +1\}^2$上的实值函数，每个这样的 f 都有唯一的展开

$$f(u, v) = \hat{f}_{\varnothing} + \hat{f}_U\, u + \hat{f}_V\, v + \hat{f}_{UV}\, uv.$$

注意^ $f_{UV} = \frac{1}{4}\mathbb{E}[UV]$。写作 $\alpha = \mathbb{P}(A)$和 $\beta = \mathbb{P}(B)$，我们得到

$$\mathbb{E}[UV] = 4p - 2\alpha - 2\beta + 1. = (2\alpha - 1)(2\beta - 1) + 4\Delta,$$

这是关于 $\mathsf{Cov}(X,\ Y)\ = \Delta$ 的仿射函数，其中边缘分布保持不变　　.

分离出边缘项后，恰好剩下一个二阶交互系数，其符号与 Δ的符号相同。这一视角在布尔函数分析和离散立方体 [14, Ch. 1].ODonnell 中很自然。

3.5　希尔伯特空间与典型相关视角

协方差证明可以进一步深化为关于 L^2上的条件期望算子的论述。

定义中心化标准化指示变量

$$u_A \coloneqq \frac{\mathbf{1}_A - \mathbb{P}(A)}{\sqrt{\mathbb{P}(A)\,\mathbb{P}(\neg A)}}, \qquad u_B \coloneqq \frac{\mathbf{1}_B - \mathbb{P}(B)}{\sqrt{\mathbb{P}(B)\,\mathbb{P}(\neg B)}}.$$

它们的均值均为零，方差均为一。 B-可测的零均值随机变量构成的空间是一维的，由 u_B 张成。因此， $\mathbb{E}[u_A \mid B]$ 是 B- 可测且均值为零的，必定是 u_B 的标量倍数。记为

$$\mathbb{E}[u_A \mid B] = \rho\, u_B.$$

与 u_B 作内积，得到 $\rho = \mathbb{E}[u_A\, u_B]$，即典型相关系数。完全相同地交换 A 和 B 后，可得

$$\mathbb{E}[u_B \mid A] = \rho\, u_A.$$

现在， u_B 在 B 上为正，在 $\neg B$上为负。因此， $\rho > 0$ 当且仅当 $\mathbb{E}[u_A \mid B] > \mathbb{E}[u_A \mid \neg B]$。由于 u_A 是 $\mathbf{1}_A$的正仿射函数，这等价于 $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$。根据对称恒等式 $\mathbb{E}[u_B \mid A] = \rho\, u_A$，同一条件 $\rho > 0$ 等价于$\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$。

因此，这两个条件期望算子是两个一维零均值希尔伯特空间之间的伴随算子，所以存在一个带符号的标量——典型相关 ρ——同时决定两个方向。在二元情形下， ρ就是皮尔逊相关系数

$$\phi = \frac{\Delta}{\sqrt{\mathbb{P}(A)\,\mathbb{P}(\neg A)\,\mathbb{P}(B)\,\mathbb{P}(\neg B)}},$$

但希尔伯特空间表述可推广到任意有限划分（此时典型相关构成一个向量，而且简单的对称性可能不再成立）。

备注3.3（正交投影）。恒等式 $\mathbb{E}[u_A \mid B] = \rho\, u_B$ 也可以通过将中心化指示变量 $\mathbf{1}_A - \mathbb{P}(A)$投影到 $L^2(\Omega)$中由1和 $\mathbf{1}_B$张成的二维子空间上推导出来。投影系数为

$$\frac{\langle \mathbf{1}_A - \mathbb{P}(A),\, \mathbf{1}_B - \mathbb{P}(B) \rangle}{\|\mathbf{1}_B - \mathbb{P}(B)\|^2} = \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(Y)},$$

即回归斜率 $\beta_{X|Y}$。交换 A 和 B 会交换同一个对称内积 $\langle \cdot, \cdot \rangle$中的两个中心化指示变量，因此符号保持不变。

4 耦合参数证明

4.1 视角

本节为定理 1.1 提供结构性解释。边缘分布固定后，一个2×2 表格只有一个自由度，因此两个条件概率差必然由同一个交互坐标决定。

4.2 证明

设 $\alpha := \mathbb{P}(A)$，且 $\beta := \mathbb{P}(B)$。具有这些边缘分布的每个 2×2 分布都由$p = \mathbb{P}(A \wedge B)$确定；其余单元格则由边缘约束确定，

$$q = \alpha - p, \qquad r = \beta - p, \qquad s = 1 - \alpha - \beta + p.$$

定义耦合参数

$$t := p - \alpha\beta,$$

它衡量 $\mathbb{P}(A \wedge B)$ 超出其独立性情况下的值 $\alpha\beta$ 的部分。因此，独立时 $t = 0$ ，正关联时 $t > 0$ ，负关联时 $t < 0$ 。

定理 *1.1* 的证明，耦合参数版本。 用 t 表示这两个条件概率差。由于 $p+r=\beta$ 和 $q+s=1-\beta$，我们有

$$\mathbb{P}(A \mid B)-\mathbb{P}(A \mid \neg B)=\frac{p}{\beta}-\frac{q}{1-\beta}=\frac{p(1-\beta)-q\beta}{\beta(1-\beta)}.$$

将 $p=\alpha\beta+t$ 和 $q=\alpha(1-\beta)-t$ 代入分子，得到

$$(\alpha\beta+t)(1-\beta)-(\alpha(1-\beta)-t)\beta=t(1-\beta)+t\beta=t.$$

因此，我们有

$$\mathbb{P}(A \mid B)-\mathbb{P}(A \mid \neg B)=\frac{t}{\beta(1-\beta)}.$$

同理，由于 $p+q=\alpha$ 和 $r+s=1-\alpha$，我们有

$$\mathbb{P}(B \mid A)-\mathbb{P}(B \mid \neg A)=\frac{p}{\alpha}-\frac{r}{1-\alpha}=\frac{t}{\alpha(1-\alpha)}.$$

两者都是 t 除以严格正分母所得的商，因此两者的符号都与 t. □

这个证明将定理 1.1 表述为关于 2×2 列联表几何的一个结论——边缘分布固定后，只有一个自由度，而且所有关于正关联的方向性概念都必须在独立点 [7]处变号。从离散 Copula 视角来看，同一个族可视为具有固定边缘分布、且仅剩一个依赖坐标 [9]的二元伯努利族。

4.3 归一化交互参数

通过如下写法，我们得到一个方便的归一化交互坐标：

$$\mathbb{P}(A \wedge B)=\alpha\beta+\theta\,\alpha(1-\alpha)\beta(1-\beta)$$

其中 θ 为标量。其余单元格则由边缘约束确定。直接计算可得：

$$\mathbb{P}(A \mid B)-\mathbb{P}(A \mid \neg B)=\theta\,\alpha(1-\alpha),\qquad \mathbb{P}(B \mid A)-\mathbb{P}(B \mid \neg A)=\theta\,\beta(1-\beta).$$

因此，两个条件概率差的符号都与 θ 相同。在二元伯努利情形下，这是对依赖结构进行重新参数化的一种方便方式，只需一个参数；在边缘分布固定时，它与 Pearson ϕ 系数仅相差一个正比例因子 $\sqrt{\alpha(1-\alpha)\beta(1-\beta)}$。事实上，我们有：

$$\phi=\frac{\Delta}{\sqrt{\alpha(1-\alpha)\beta(1-\beta)}}=\theta\sqrt{\alpha(1-\alpha)\beta(1-\beta)}$$

[9, 16]。

4.4 微分观点

耦合参数证明有一种自然的微分细化。对于每个可容许值 t，令 $\mathbb{P}_t$ 表示具有固定边缘分布 $\mathbb{P}_t(A) = \alpha$ 和$\mathbb{P}_t(B) = \beta$ 且耦合参数为 t 的 $(A,\ B)$ 的唯一联合分布，即:

$$\mathbb{P}_t(A \wedge B) = \alpha\beta + t, \qquad \mathbb{P}_t(A \wedge \neg B) = \alpha(1-\beta) - t,$$

$$\mathbb{P}_t(\neg A \wedge B) = \beta(1-\alpha) - t, \qquad \mathbb{P}_t(\neg A \wedge \neg B) = (1-\alpha)(1-\beta) + t.$$

这里，“可容许”确切地意味着这四个量均非负，即

$$-\min\{\alpha\beta, (1-\alpha)(1-\beta)\} \le t \le \min\{\alpha(1-\beta), (1-\alpha)\beta\}.$$

现在定义

$$F(t) \coloneqq \mathbb{P}_t(A \mid B) - \mathbb{P}_t(A \mid \neg B) = \frac{t}{\beta(1-\beta)},$$
$$G(t) \coloneqq \mathbb{P}_t(B \mid A) - \mathbb{P}_t(B \mid \neg A) = \frac{t}{\alpha(1-\alpha)}.$$

那么，F 和 G关于 t 都是线性的，且斜率严格为正

$$F'(t) = \frac{1}{\beta(1-\beta)} > 0, \qquad G'(t) = \frac{1}{\alpha(1-\alpha)} > 0,$$

并且 $F(0) = G(0) = 0$。由于 F 和 G 均严格递增且经过原点，我们有 $\operatorname{sgn} F(t) = \operatorname{sgn} t = \operatorname{sgn} G(t)$，定理 1.1 得证。

4.5 一致性与不一致性

从联合分布中独立抽取两次，得到抽样对 $(X,\ Y)$ 和 $(X',\ Y')$，其中 $X = \mathbf{1}_A, Y = \mathbf{1}_B$。若一个样本为 $(1,\ 1)$，另一个为 $(0,\ 0)$，则称该抽样对一致；若一个样本为 $(1,\ 0)$，另一个为 $(0,\ 1)$，则称其不一致。于是

$$\mathbb{P}(\text{concordant}) = 2ps, \qquad \mathbb{P}(\text{discordant}) = 2qr.$$

因此，$\Delta > 0$ 当且仅当一致的抽样对比不一致的抽样对更有可能出现。显然，这一条件在交换两个坐标时保持不变，由此又得到了定理 1.1 的一个证明。B 是 A 的证据，当且仅当在两次独立抽样中，同向结果比交叉结果更有可能出现。这让人联想到肯德尔的 τ [12]。

与耦合参数的联系显而易见，因为

$$\mathbb{E}[(X - X')(Y - Y')] = 2\operatorname{Cov}(X, Y) = 2\Delta,$$

而乘积 $(X - X')(Y - Y')$ 在一致对上等于 $+1$，在不一致对上等于 -1，其他情况下为 0。

4.6 传输与不匹配成本

对于边缘分布固定为 α 和 β 的伯努利变量 $X = \mathbf{1}_A$ 和 $Y = \mathbf{1}_B$，汉明不匹配概率为

$$\mathbb{P}(X \neq Y) = q + r. \tag{3}$$

在独立性条件下，(3) 等于 $\alpha(1-\beta)+(1-\alpha)\beta$。正关联（$\Delta > 0$）等价于实际耦合的不匹配汉明距离小于独立性耦合，即:

$$q + r < \alpha(1-\beta) + (1-\alpha)\beta.$$

展开后，可化简为 $p > \alpha\beta$，即 $t > 0$。因此，证据为正等同于说，联合分布在离散最优传输的意义上，是一种优于独立性耦合的耦合。

5 优势比证明

5.1 视角

本节从统计学和贝叶斯推断所采用的优势尺度来解读定理 1.1。这里的关键不变量是优势比 ps/qr，它相对于 1 的大小关系决定了两种条件比较的方向，且对二者一视同仁。

5.2 证明

定理 1.1 的证明，优势比版本。 函数 $\pi \mapsto \pi/(1-\pi)$ 在单位区间 $(0,\ 1)$ 上严格递增。因此，当且仅当给定 $\neg B$ 时的优势小于给定 A 时的优势，$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$。给定 A 和 B 时的优势为

$$\frac{\mathbb{P}(A \mid B)}{\mathbb{P}(\neg A \mid B)} = \frac{p}{r},$$

给定 $\neg B$ 时的优势为 q/s。不等式变为

$$\frac{p}{r} > \frac{q}{s} \iff ps > qr \iff \Delta > 0.$$

同样，$\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$ 可转化为 $p/q > r/s$，这又等价于 $ps > qr$。两者都等价于优势比 OR $:= ps/qr$ 大于 1 [1]。 □

备注 5.1（历史注记）。 作为关联度量的优势比至少可以追溯到尤尔 [17]。然而，在本节中，[1]是关于 2×2 表格中正关联、优势比与逻辑回归系数之间等价关系更直接的技术参考。

5.3 贝叶斯证据互惠性

赔率形式的贝叶斯定理 [10, 第 1] 章指出，观察到 B 后，A 的后验赔率与先验赔率之间的关系为

$$\underbrace{\frac{\mathbb{P}(A \mid B)}{\mathbb{P}(\neg A \mid B)}}_{\text{posterior odds}} = \underbrace{\frac{\mathbb{P}(A)}{\mathbb{P}(\neg A)}}_{\text{prior odds}} \cdot \underbrace{\frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B \mid \neg A)}}_{\text{Bayes factor}}.$$

贝叶斯因子大于 1，当且仅当 $\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$，这意味着观察到 B 会提高 A 的赔率。这等价于 OR > 1，其中 A 和 B 是对称的。定理 1.1 的贝叶斯解读是，证据支持具有互惠性——B 是支持 A 的证据，当且仅当 A 是支持 B [6, 4] 的证据。

备注 5.2。这种互易性适用于二元划分。一旦转向更细的划分或条件分层，就可能出现相关的逆转；参见辛普森悖论 [2]。

5.4 伊辛相互作用与指数族

对数优势比 $\log(ps/qr)$ 是将 2×2 表格视为对数线性模型或指数族模型时的自然交互参数。将联合概率质量函数写成吉布斯形式：

$$\mathbb{P}(U = u, V = v) = \frac{1}{Z} \exp\bigl(\alpha\, u + \beta\, v + \gamma\, uv\bigr), \qquad u, v \in \{-1, +1\},$$

其中 U，V是 A，B的 ± 1 编码。直接计算可得：

$$\gamma = \tfrac{1}{4} \log(ps/qr).$$

因此，唯一交互参数 γ 的符号与对数优势比的符号相同，因而也与 Δ的符号相同。用统计力学的术语来说，$\gamma > 0$ 是**铁磁性的**（有利于排列一致的构型），$\gamma < 0$ 是**反铁磁性的**，而 $\gamma = 0$表示独立性。关于交互作用与对数优势比之间具体对应关系的 2×2 说明，见 [1]；关于更广泛的对数线性和图模型视角，另见 [13]。

5.5 单调似然比

对于二元观测 $Y \in \{0, 1\}$，定义似然比

$$\Lambda(v) := \frac{\mathbb{P}(V{=}v \mid A)}{\mathbb{P}(V{=}v \mid \neg A)}.$$

那么 $\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$ 等价于 $\Lambda(1) > \Lambda(0)$，因为 $\Lambda(1) = {}^{p}/_{(p+q)}\ {}^{r}/_{(r+s)}$ 且 $\Lambda(0) = {}^{q}/_{(p+q)}\ {}^{s}/_{(r+s)}$，所以 $\Lambda(1)/\Lambda(0) = ps/qr$ 是优势比。根据赔率形式的贝叶斯定理，$\Lambda(1) > \Lambda(0)$ 意味着，观测到 $Y{=}1$ 会使 A 的优势乘上一个比观测到 $Y{=}0$ 时更大的因子，也就是 $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$。这一带有内曼–皮尔逊风格的论证 [11, Ch. 1] 在技术上类似于贝叶斯因子的证明；不过，在概念上二者有所区别，因为它将结果表述为族 $\{\mathbb{P}(\cdot \mid A)\ \mathbb{P}(\cdot \mid \neg A)\}$ 的单调似然比性质。

5.6 逻辑回归

在一个 2×2 表格中，对 $\mathbf{1}_A$ 关于 $\mathbf{1}_B$ 进行饱和逻辑回归，得到

$$\operatorname{logit} \mathbb{P}\left(X{=}1 \mid Y{=}b\right) = \hat{\beta}_0 + \hat{\beta}_1\, b, \qquad b \in \{0, 1\},$$

ˆ其中 $\beta_1 = \log(ps/qr)$ [1, 见 2×2 表格、优势比和逻辑回归]。对 $\mathbf{1}_B$ 关于 $\mathbf{1}_A$ 进行ˆ 反向回归，得到相同的系数，因为 β_1 在两种情况下都是对数优势比。因此

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff \hat{\beta}_1 > 0 \iff \mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A).$$

这是用广义线性模型的语言表述的优势比证明。

6 信息论证明

6.1 视角

本节用信息论术语重新表述定理 1.1。关键在于，正向证据支持等价于正点互信息，而正点互信息关于 A 和 B 的对称性使定理 1.1 的互易性显而易见。

6.2 证明

定义 6.1 （点互信息）。 结果对（A， B）的点互信息 （PMI）为 [3, 第 2 章，尤其是 §2.3]

$$\operatorname{pmi}(A;B) := \log \frac{\mathbb{P}(A \wedge B)}{\mathbb{P}(A)\,\mathbb{P}(B)} = \log \frac{\mathbb{P}(A \mid B)}{\mathbb{P}(A)} = \log \frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B)}.$$

根据贝叶斯定理，这三个表达式相等；第一个表达式清楚地体现了逐点互信息 pmi $(A; B) =$ pmi $(B;\ A)$ 的对称性。

引理 6.2. </style}当且仅当 pmi ($A; B) > 0$时，有 $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$ *</style}*。

证明. 根据全概率公式，我们计算得

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid \neg B)\,\mathbb{P}(\neg B).$$

由于 $\mathbb{P}(B)$ 和 $\mathbb{P}(\neg B)$ 严格为正且和为 1，因此边缘概率 $\mathbb{P}(A)$ 是 $\mathbb{P}(A \mid B)$ 和 $\mathbb{P}(A \mid \neg B)$ 的严格凸组合。除非两者相等，否则它严格介于两者之间。因此

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff \mathbb{P}(A \mid B) > \mathbb{P}(A).$$

由于 log 严格递增， $\mathbb{P}(A \mid B) > \mathbb{P}(A) \iff$ 逐点互信息 $(A; B) > 0$。 □

*Pro*定理 *1.1* 的信息论版本。根据引理 6.2，我们有 e

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff \operatorname{pmi}(A;B) > 0.$$

根据对称性，

$$\operatorname{pmi}(A;B) > 0 \iff \operatorname{pmi}(B;A) > 0.$$

由引理 6.2 交换 A 和 B 可得，

$$\operatorname{pmi}(B;A) > 0 \iff \mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A).$$

□

注 6.3（卡尔纳普的概率比）。这个证明与贝叶斯确证文献所称的概率比规则 [6, 4], 密切相关，即 $\mathbb{P}(A \mid B)/\mathbb{P}(A) = \mathbb{P}(B \mid A)/\mathbb{P}(B)$。这是去掉对数后的点互信息。我们不将其视为一个独立的证明，但它值得一提，因为它是这一方法的历史先驱。

6.3 惊异度降低

A 的惊异度为 $-\log \mathbb{P}(A)$；学习到 B 后，它变为 $-\log \mathbb{P}(A \mid B)$。惊异度的降低量为

$$\log \mathbb{P}(A \mid B) - \log \mathbb{P}(A) = \operatorname{pmi}(A; B).$$

正值意味着，在学习到 B后，A 变得不那么令人意外。根据逐点互信息的对称性，同一个量也衡量了 A 使 B的惊异度降低了多少。证据互惠性就是惊异度降低的互惠性。

6.4 互信息与依赖方向

普通互信息

$$I(A; B) \coloneqq \mathbb{E}[\operatorname{pmi}(A; B)]$$

总是非负的 [3, 第 2], 章，并且当且仅当 A 和 B 相互独立时为零。它衡量依赖的程度，但不反映依赖的方向：正相关和负相关的二元变量对可能具有相同的互信息。

关联方向则体现在四个原子结果对的点互信息层面上。用我们的记号表示,

$$\operatorname{pmi}(A; B) = \log \frac{p}{(p+q)(p+r)}, \qquad \operatorname{pmi}(\neg A; \neg B) = \log \frac{s}{(r+s)(q+s)},$$

而

$$\operatorname{pmi}(A; \neg B) = \log \frac{q}{(p+q)(q+s)}, \qquad \operatorname{pmi}(\neg A; B) = \log \frac{r}{(r+s)(p+r)}.$$

直接计算可得

$$\operatorname{sgn} \operatorname{pmi}(A; B) = \operatorname{sgn} \operatorname{pmi}(\neg A; \neg B) = \operatorname{sgn} \Delta$$

and

$$\operatorname{sgn} \operatorname{pmi}(A; \neg B) = \operatorname{sgn} \operatorname{pmi}(\neg A; B) = -\operatorname{sgn} \Delta,$$

其中 $\Delta = ps - qr$同往常一样。因此，对于二元变量，两个一致原子 $A \wedge B$和$\neg A \wedge \neg B$ 具有相同的点互信息符号，两个不一致原子 $A \wedge \neg B$和 $\neg A \wedge B$ 具有相反的符号，而同一个标量 Δ 支配着它们全部。

7 组合证明

7.1 视角

本节给出定理 1.1 的一个初等计数证明。它不对概率或概率的变换进行运算，而是比较有利和不利的有序对，并证明同一个乘积不等式在两个方向上都会出现。

7.2 证明

假设概率空间是有限且均匀的，并设有子集

$$\mathcal{P} \coloneqq A \cap B, \quad \mathcal{Q} \coloneqq A \cap \neg B, \quad \mathcal{R} \coloneqq \neg A \cap B, \quad \mathcal{S} \coloneqq \neg A \cap \neg B,$$

其基数分别为 $|\mathcal{P}| = n_p$、$|\mathcal{Q}| = n_q$、$|\mathcal{R}| = n_r$、$|\mathcal{S}| = n_s$。

定理 *1.1* 的证明，组合版本。为了比较 $\mathbb{P}(A \mid B)$ 和 $\mathbb{P}(A \mid \neg B)$，考虑满足 $u \in B$ 和 $v \in \neg B$ 的有序对 (u, v)。这样的有序对共有 $(n_p + n_r)(n_q + n_s)$ 个。若满足 $u \in A$ 和 $v \notin A$，则称该有序对为 A-有利的；若满足 $u \notin A$ 和 $v \in A$，则称其为 A-不利的。A-有利的有序对恰好是满足 $u \in \mathcal{P}$ 和 $v \in \mathcal{S}$ 的那些，因此共有 $n_p n_s$ 个。A-不利的有序对满足 $u \in \mathcal{R}$ 和 $v \in \mathcal{Q}$，因此共有 $n_r n_q$ 个。因此

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff n_p n_s > n_r n_q.$$

现在交换 A 和 B 的角色。为了比较 $\mathbb{P}(B \mid A)$ 和 $\mathbb{P}(B \mid \neg A)$，考虑满足 $u \in A$ 和 $v \in \neg A$ 的有序对 (u, v)。B-有利的有序对满足 $u \in \mathcal{P}$ 和 $v \in \mathcal{S}$，仍由 $n_p n_s$ 计数。B-不利的有序对满足 $u \in \mathcal{Q}$ 和 $v \in \mathcal{R}$，由 $n_q n_r$ 计数。因此

$$\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A) \iff n_p n_s > n_q n_r.$$

这两个不等式完全相同。 □

7.3 推广到任意概率空间

组合证明可直接用于有限均匀空间。对于有理概率，可通过消去分母将问题化为有限均匀情形。对于所有条目都严格为正的实值表格，这两个条件概率差

$$F(p, q, r, s) \coloneqq \mathbb{P}(A \mid B) - \mathbb{P}(A \mid \neg B) \text{ and}$$
$$G(p, q, r, s) \coloneqq \mathbb{P}(B \mid A) - \mathbb{P}(B \mid \neg A)$$

是开单纯形 $\{(p, q, r, s) \in (0, 1)^4 \mid p + q + r + s = 1\}$上的连续函数。在有理数表的稠密子集上，$\operatorname{sgn} F = \operatorname{sgn} G$，且这两个函数都是连续的，因此符号相等这一性质可推广到整个单纯形。

7.4 随机配对解释

组合证明可以改写为概率形式。独立地从 B 中均匀抽取一点u，再从 $\neg B$ 中均匀抽取一点 v。于是 $\mathbb{P}(u \in A) = \mathbb{P}(A \mid B)$ 且$\mathbb{P}(v \in A) = \mathbb{P}(A \mid \neg B)$，因此 $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$ 表明，B点满足A 的可能性高于 $\neg B$点。满足条件的点对数 $n_p n_s$ 和不满足条件的点对数 $n_r n_q$，分别是以下事件的联合概率的缩放：“u满足 A而 v 不满足”以及“u不满足 A 而 v 满足”——这是一种自然的锦标赛式比较。

8 对数势证明

8.1 视角

本节用 $\{0,1\}^2$ 上的离散势来重新表述定理 1.1。关键在于，两个条件不等式都等价于 log P 的同一个混合离散导数为正，因此，这种对称性源自离散微分的可交换性。

8.2 证明

令 $g(a,b) \coloneqq \log \mathbb{P}(A{=}a,\, B{=}b)$ 表示 $(a,b) \in \{0,1\}^2$，因此

$$g(1,1) = \log p, \quad g(1,0) = \log q, \quad g(0,1) = \log r, \quad g(0,0) = \log s.$$

定义水平对数优势比场

$$H(b) \coloneqq g(1,b) - g(0,b) = \log \frac{\mathbb{P}(A{=}1 \mid B{=}b)}{\mathbb{P}(A{=}0 \mid B{=}b)},$$

即，$H(b)$ 是 B 在 b 水平处 A 的对数优势比。由于逻辑函数 : $\underline{\pi \mapsto \log\left(\frac{\pi}{1-\pi}\right)}$ 严格递增，

$$\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B) \iff H(1) > H(0).$$

定义垂直对数优势比场

$$V(a) \coloneqq g(a,1) - g(a,0) = \log \frac{\mathbb{P}(B{=}1 \mid A{=}a)}{\mathbb{P}(B{=}0 \mid A{=}a)}.$$

Then

$$\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A) \iff V(1) > V(0).$$

定理 *1.1* 的证明，对数势版本。.....的混合离散导数 g 为

$$\begin{aligned} H(1) - H(0) &= \big[g(1,1) - g(0,1)\big] - \big[g(1,0) - g(0,0)\big] \\ &= g(1,1) + g(0,0) - g(1,0) - g(0,1). \end{aligned}$$

同一个量也可由 V 得到，

$$\begin{aligned} V(1) - V(0) &= \big[g(1,1) - g(1,0)\big] - \big[g(0,1) - g(0,0)\big] \\ &= g(1,1) + g(0,0) - g(1,0) - g(0,1). \end{aligned}$$

混合离散导数 $\partial_{AB} g \coloneqq g(1,\,1) + g(0,\,0) - g(1,\,0) - g(0,\,1)$ 不依赖于差分的求取顺序。（这相当于 $\partial_a \partial_b = \partial_b \partial_a$ 的离散版本。）由于逻辑函数严格递增，$H(1) > H(0)$ 和 $V(1) > V(0)$ 都等价于 $\partial_{AB} g > 0$。 □

注意，$\partial_{AB} g = \log(ps/qr)$，因此 $\partial_{AB} g > 0$ 当且仅当 OR > 1 当且仅当 $\Delta > 0$。但这一证明的核心机制不同于优势比证明——不变量并非表现为一个比值，而是势的混合增量；对称性则源自离散微分的可交换性，而非对优势进行任何代数运算。

8.3 全正性与 FKG不等式

条件 $\partial_{AB} g \geq 0$——等价地说， $ps \geq qr$——是二阶全正性（TP_2） [11, 第 1], 章中的 2×2 情形，或者说是对数超模性。若非负矩阵的每个2×2 子式都非负，则该矩阵为 TP_2 。对于 2×2 表格，只有一个子式，因此 TP_2 可归结为$ps \geq qr$。

对数超模性会带来显著的推论。著名的 ***FKG*不等式** [8]指出，如果有限分配格上的联合分布具有对数超模性，那么任意两个递增事件都正相关。在 2×2情形下，这恰好就是定理 1.1。全正性与对数超模性在矩阵转置下都是不变量，因此，定理 1.1 中的对称性是一个普遍现象的特例。

注 8.1。 在连续情形中，联合密度$f(a, b)$的对数超模性对应条件是 $\partial^2 \log\ f/\partial a\, \partial b \geq 0$。混合离散导数 $\partial_{AB} g$是这一条件的离散版本。

8.4 随机序

将两个条件分布 $\mu_A := \mathcal{L}(B \mid A)$和 $\mu_{\neg A} := \mathcal{L}(B \mid \neg A)$视为有序二点空间 $\{0, 1\}$上的概率测度。那么， $\mathbb{P}(B \mid A) > \mathbb{P}(B \mid \neg A)$恰好表示$\mu_A \succ_{\mathrm{st}} \mu_{\neg A}$ 通常随机序意义下的关系，见第 [15, 章 1]。同理， $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid \neg B)$表示交换坐标后也成立类似的随机支配。对于 2×2 表格，两个条件分布的带符号差只有一个自由度，因此按行随机序与按列随机序等价。对于满足 $\min(m, n) \geq 3$ 的 $m \times n$ 表格，这些概念可能不再一致，这就是为什么在高维中简单的对称性会失效。

8.5 对数线性模型

混合增量 $\partial_{AB} g = \log(ps/qr)$ 恰好是 2×2 表格的饱和对数线性模型中的交互项。用对数线性形式表示，联合质量函数可写为

$$\log \mathbb{P}(A{=}a,\, B{=}b) = \mu + \lambda_A^{(a)} + \lambda_B^{(b)} + \lambda_{AB}^{(a,b)},$$

并遵循通常的可识别性约定，即交互项对每个指标求和均为零。在 2×2 情形下，交互作用有一个自由参数；按照标准的效应编码约定，该参数为 $\tfrac{1}{4}\log(ps/qr)$。因此，对数势证明将定理 1.1 归结为饱和对数线性模型 [1] 中唯一交互参数为正。

9 结论

我们给出了定理 1.1 的七种不同证明。

1. **代数证明** 通过交叉相乘处理条件概率分式。

2. **协方差证明**通过 $\mathbb{P}(A \mid B) - \mathbb{P}(A) = \mathrm{Cov}\,/\mathbb{P}(B)$ 将每个条件概率与边缘概率联系起来，并利用 $\mathrm{Cov}\,(X,\ Y) = \mathrm{Cov}\,(Y,\ X)$。

3. **耦合参数证明**将两个条件概率差表示为$t/$(正数)，其中边缘分布固定后，t 是唯一的自由参数。4. **优势比证明**转而考察优势，从而化简为 OR >1。5. **信息论证明**利用定义对称性 pmi $(A;\ B)$ =pmi $(B;\ A)$ 和全概率公式，从未提及 Δ。6. **组合证明**通过两种不同的分解方式，分别计算有利和不利的有序对。7. **对数势证明**将两个不等式都归结为 log P 的混合离散导数为正，并依赖离散微分的交换性。

这些证明源于实质上不同的视角。代数证明、协方差证明和优势比证明分别通过不同途径得到标量 $\Delta = ps - qr$ （或其单调变换）——直接进行分式运算、以边缘概率为锚点，以及转而考察优势。耦合参数证明则改为研究 $t = p - \mathbb{P}(A)\mathbb{P}(B)$：在边缘分布固定后，它是唯一的交互坐标。信息论证明则完全避开了 Δ。组合证明通过计数对象来证明。对数势证明采用不同的尺度——考察对数 P而非 P——并从离散算子的可交换性推导出对称性，而不是通过对分式进行代数运算得出。

围绕这些证明，我们构建了多种视角交织。下表总结了我们关注的主要关注量，以及它们为正时所代表的含义。

视角	量	正值意味着……
线性代数	$\det M = \Delta$	秩大于 1
几何	行向量的有向面积	有利倾斜
概率	$\mathrm{Cov}(\mathbf{1}_A,\ \mathbf{1}_B)$	正相关性
回归	斜率 $\beta_{X\mid Y}$	递增条件均值
沃尔什–傅里叶	交互作用 $\hat{f}_{UV}$	正交互作用
希尔伯特空间	典型相关 ρ	对齐投影
耦合	$t = p - \mathbb{P}(A)\mathbb{P}(B)$	上述独立性
一致性	$\mathbb{P}(\text{一致性}) > \mathbb{P}(\text{不一致性})$	更倾向于对齐对
传输	$\mathbb{P}(X \neq Y)$	比独立情况的不匹配更少
统计学	$\log(ps/qr)$	优势比 > 1
贝叶斯	贝叶斯因子 > 1	B 支持 A
伊辛模型	交互作用 γ	铁磁耦合
MLR	$\Lambda(1)/\Lambda(0)$	单调似然比
逻辑斯蒂	$\hat{\beta}_1 > 0$	正逻辑斯蒂斜率
信息	逐点互信息 $(A;B)$	正局部证据
组合	$n_p n_s - n_q n_r$	有利有序对占主导
对数势	$\partial_{AB} \log P$	正混合增量
全正性	TP_2 子式	正关联
随机序	$\mu_A \succ_{\mathrm{st}} \mu_{\neg A}$	逐行占优
对数线性	$\lambda_{AB} > 0$	正交互项

在 2×2 情形下，所有标量关联度量都归结为一个参数，因为可能的表格构成一个单参数族。

命题 9.1。四个元素 p, q, r, s 满足三个独立的线性约束（$p+q = \mathbb{P}(A)$、$p + r = \mathbb{P}(B)$、$p + q + r + s = 1$），因此 p 是唯一的自由参数。因此，任何仅取决于联合分布的二元关联有符号测度，都是 p的函数，或者等价地是 Δ的函数，也等价地是耦合参数 t的函数。

这种一维性解释了为什么上表中的全部十九种视角对关联的符号看法一致：它们是同一个一维族的不同坐标系。对于 $m \times n$ 个最小值为m、$n \geq 3$的表格，交互结构会变成多维的，单个标量不再能支配所有合理的关联概念；例如，希尔伯特空间典型相关会变成一个向量，优势比会推广为局部优势比矩阵，而按行随机序与按列随机序可能会得出不同结论。二元情形很特殊，因为一个数就能说明全部情况。

参考文献

[1] A. Agresti，《分类数据分析》，第2版，Wiley，2002年。[2] C. R. Blyth，《论辛普森悖论与确定性原则》，《美国统计学会会刊》67（1972），364–366。[3] T. M. Cover 和 J. A. Thomas，《信息论基础》，第2版，Wiley，2006年。[4] V. Crupi、K. Tentori 和 M. Gonzalez，《论证据支持的贝叶斯度量：理论与实证问题》，《科学哲学》74（2007），229–252。[5] M. Drton、B. Sturmfels 和 S. Sullivant，《代数统计学讲义》，Birkh¨auser，巴塞尔，2009年。[6] E. Eells 和 B. Fitelson，《证据支持中的对称性与非对称性》，《哲学研究》107（2002），129–142。[7] S. E. Fienberg 和 J. P. Gilbert，《二乘二列联表的几何》，《美国统计学会会刊》65（1970），694–701。[8] C. M. Fortuin、P. W. Kasteleyn 和 J. Ginibre，《若干偏序集上的相关不等式》，《数学物理通讯》22（1971），89–103。[9] G. Geenens，《离散随机向量的Copula建模》，《依赖建模》8（2020），417–440。[10] I. J. Good，《概率与证据权衡》，Charles Griffin，伦敦，1950年。[11] S. Karlin，《全正性》，第I卷，斯坦福大学出版社，1968年。[12] M. G. Kendall，《一种新的秩相关度量》，Biometrika 30（1938），81–93。[13] S. L. Lauritzen，《图模型》，牛津大学出版社，1996年。[14] R. O’Donnell，《布尔函数分析》，剑桥大学出版社，2014年。[15] M. Shaked 和 J. G. Shanthikumar，《随机序》，Springer，纽约，2007年。[16] M. J. Warrens，《关于 2×2 表的关联系数及不依赖边缘分布的性质》，Psychometrika 73（2008），777–789。[17] G. U. Yule，《统计学中的属性关联：并以儿童社会等材料为例》，《伦敦皇家学会哲学汇刊：A辑》194（1900），257–319。[18] G. U. Yule，《论测量两个属性间关联的方法》，《英国皇家统计学会会刊》75（1912），579–652。