arXiv:2601.10641v1 [stat.ME] 15 Jan 2026

调整后的相似性度量与期望违背

William L. Lippitt、Edward J. Bedrick 和 Nichole E. Carlson

摘要

调整后的相似性度量，例如用于评定者间信度的 Cohen κ，以及用于比较聚类算法的调整Rand指数，是比较离散标注的重要工具。这些度量旨在满足以下性质：在零分布下的期望为 0，且在相似性达到最大时取最大值 1，以便于解释。出于历史和分析方面的原因，这些度量经常根据置换分布进行调整。目前，人们重新开始关注其他更适合具体情境的零模型，例如允许识别出的聚类数量随机变化的聚类集成。本研究有两个目的：（1）将对调整算子的研究推广至一般零模型，并推广至一种更一般的程序，该程序将统计标准化作为特例；（2）确定调整算子能够产生预期性质的充分条件，其中充分条件与观测数据是否以及如何纳入零分布有关。我们展示了违反充分条件可能导致严重失效，例如，传统调整可能产生非正值度量，而非均值为 0 的度量；统计标准化则可能产生恒为 0 的度量。

关键词：相似性指数、机会校正、Cohen κ、调整Rand指数、统计标准化

1 引言

假设我们有两个要进行比较的分类变量，例如使用两种不同评估技术进行的诊断评估，或真实标注与通过无监督聚类或预测建模生成的标签。我们可以使用一个指标，也就是一个由这两个变量的联合经验分布定义的函数，来量化它们的相似性。文献中针对不同情境提出了大量指标，例如 Cohen 的 κ [Cohen，用于 1960] 评估取值范围相同的变量之间的一致性；Rand 指数的各种变体 [Rand, 1971, Hubert and Arabie,1985] 以及互信息 [Shannon, 1948, Romano 等人， 2016, 2014] 用于变量划分簇成员关系的情形；还有 Goodman 和 Kruskal 的 γ [Goodman and Kruskal，1979]用于衡量有序变量之间的秩相关。研究者已致力于理解各指标之间的关系及其性质 [Albatineh 等，2006, Meilˇa, 2005, Warrens，2013, Brusco et al., 2021, Arinik et al., 2021]以期识别冗余和不理想的行为，并将应用情境与适当的指标相匹配。

长期以来，人们一直关注如何对指数进行调整，以消除“随机机会”的影响。具体做法是，先减去相对于某个零模型的期望值，该零模型用于描述由“随机机会”导致的预期相似性，然后再除以最大值 [Scott，1955年，Cohen，1960年，Hubert 和 Arabie，1985年，Krippendorff，1987年，Davenport Jr 和 El-Sanhurry，1991年，Albatineh 等，2006年，Vinh 等人，2009年，Warrens， 2013]。因此，调整后的指数旨在使未关联的比较变量对应的值接近 0，并使最大值达到最优值 1；这些性质有助于解释和比较 [Vinh 等人，2009年，Albatineh 等，2006年，Warrens， 2013]。因此，调整需要指定一个零模型和一种归一化方法。

通常将置换模型（又称超几何模型）视为随机机会的零模型 [Hubert 和 Arabie，1985年，Albatineh 等人，2006年，Vinh 等人，2009年，Romano 等人， 2016]。针对特定情境，也有人考虑其他模型 [Jeub 等人，2018年，Sundqvist 等人，2023年，Cohen，1960年，Krippendorff，1987年，Scott，1955年，Li 等人， 2024]，例如包含变量独立性的多项分布，以及允许簇数变化的聚类集成。更广泛而言，Krippendorff [1987] 阐明了不同的建模假设会如何导致对某一指标所测量内容产生截然不同的解释；Gates 和 Ahn [2017] 则表明，调整中采用的零模型会对结论产生重大影响。因此，必须根据具体情境谨慎选择零模型。

在本研究中，我们推广了关于偶然校正这一操作性质的基础研究 [Albatineh 等人，2006年，Warrens， 2013] 这些研究传统上侧重于置换模型及相关模型和特定归一化方法。我们针对任意零模型和有效的归一化选择，考察了广义调整。因此，广义调整也涵盖了 Romano 等人 [2014, 2016] 最近针对聚类考察的统计标准化这一特例。

通过对调整进行广义研究，我们作出如下贡献：

- 调整算子表现出所需性质的充分条件，例如：– 调整后的指数在零分布下的均值为 0，– 标准化后的指数在零分布下的方差也为 1，以及– 调整算子具有幂等性（无需重复调整）；
- 根据零分布是否以及如何具有数据驱动性，对充分条件进行刻画；以及
- 使用经典数据驱动零模型，展示在不满足充分条件时所需性质如何失效。

通过这种方式，我们为未来使用更一般模型对指数进行调整的方法论研究提供了具体指导：明确保证有利性质的条件，并强调在不满足这些条件时需要进行更加谨慎的理论处理。

本文的组织结构如下。我们在第 2 节正式介绍指数，并讨论相对于一般归一化和零模型调整指数的一般流程。随后，我们在第 3 节扩展先前关于调整下指数族的研究，确定有利性质的充分条件。我们给出两个违反充分条件的指数-模型对示例：其中一个在针对期望和最大值而非均值 0 进行传统调整后为非正，另一个在统计标准化后恒等于 0。第 4 节包含讨论。

2 针对机遇的总体调整

假设我们有两个分类变量要进行比较，其变量标签可能相同，也可能不同。我们可以使用一个指标来量化相似性。形式化地说：

定义 2.1 (指标 S)。 设 x 和 y 是针对相同 $k = 1,...,N$ 个观测值记录的两个分类变量，其中每个 x_k 取 I 个可能值之一，每个 y_k 取 J 个可能值之一。定义边际计数 u, v，其中 $u_i = \#\{k\colon x_k = i\}, v_j = \#\{k\colon y_k = j\}$，以及联合计数 n，其中 $n_{ij} = \#\{k\colon x_k = i, y_k = j\}$。 x 和 y 的指标 S 是一个实值函数 $S(n)$。

广义而言，指标 S 是列联表（即联合计数分布）的实值函数 n，有时针对 n 特定大小的列联表（例如，$I = J = 2$ 用于比较二元变量），有时则适用于任意大小的列联表（例如， I， $J \in \mathbb{N}$ 用于比较聚类标签）[Warrens， 2013]。在实践中，我们希望指标 S 能够以某种方式单调地衡量 x 与 y [Krippendorff， 1987] 之间相似性的强弱。下面给出两个例子，它们经过调整后可以得到许多常用的度量：

定义 2.2 (原始一致性比例 p)。 在定义 2.1的情境下，假设 x 和 y 取值于同一个由 $I = J$ 个可能值组成的指标集。 p 与 x 之间的原始一致性比例为 y： $p(n) = \sum_i \frac{n_{ii}}{N}$.

定义 2.3 (成对一致性比例 q)。 采用定义 2.1中的记号，设 w 是对某变量进行 N次观测所得的计数，该变量取值于集合 $\mathcal{R}$ (例如， $w = u$ 和 $\mathcal{R} = \{1, ..., I\}$ 或 $w = n$ 和 $\mathcal{R} = \{1, ..., I\} \times \{1, ..., J\}$)。分配到相同取值的观测值对的比例为 q： $q(w) = \frac{\sum_{r\in\mathcal{R}} \binom{w_r}{2}}{\binom{N}{2}}$。

尽管对于连续变量，许多相似性度量在不存在关联时的期望值为 0，但对于评估分类变量之间共享信息或关联的基本指数，情况通常并非如此。在 $p(n)$ 和 $q(n)$ 的情况下，值为 0 表明观测值通常不太可能是由独立变量 x 和 y随机产生的。因此，这些指数通常会通过减去期望值并除以最大值进行归一化，从而施加以下性质：0 表示从无关联变量中预期观察到的关联水平，而 1 表示可观察到的最大关联。通过这种方式，可以从 $p(n)$ 构建出更为熟悉的关联度量，例如 Cohen 的 κ [Cohen、 1960], Scott 的 π [Scott、 1955], 以及戴斯系数 [Dice， 1945] 以及从 $q(w)$ 构建出的调整Rand指数 ARI [Hubert 和 Arabie， 1985]以及用于聚类的其他成对计数度量 [Albatineh 等人， 2006] 。

2.1 anIndex 的调整

期望值和最大值的指数调整传统上是线性的，并且需要明确指定归一化常数，通常是由指数衡量的最大关联水平 S_{max} 的关联度量S，以及相对于零模型的期望值 $\mathrm{E}[S]$ $\mathcal{M}$ [Scott，1955年，Cohen，1960年，Fleiss，1975年，Hubert 和 Arabie，1985年，Krippendorff，1987年，Albatineh 等人，2006年，Warrens，2008a]。从概念上讲，我们可能会预期采用以下公式

$$AS(n) = \frac{S(n) - \mathrm{E}[S(\tilde{n})]}{S_{max} - \mathrm{E}[S(\tilde{n})]},$$

其中 $\tilde{n}$ 服从零分布，但这种表述在实践中过于僵化。有时，我们希望将观测计数 n 的某些方面纳入零分布或随机分布，例如在断言置换模型（定义 2.5）时固定观测到的边际计数 u 和 v。在定义最大值时也可能出现类似情况，此时边际计数 u 和 v 限制了可达到的最大值。因此，我们必须以非常一般的方式给出调整的定义。与以往的研究不同，我们选择明确说明零分布和最大值，并指出它们可能具有数据驱动的性质，以帮助阐明后文中指数调整出现的反直觉行为。我们用 $\amalg$ 表示不相交并，即自然互不重叠的集合之并。

定义 2.4 (调整指数 AS)。 考虑定义 2.1中的情境。令

$$\mathcal{M} = \{M^{n'} : n' \in B \subseteq \amalg_{I,J}\mathbb{N}^{I\times J}\}$$

是联合计数 $\tilde{n}$ 的一个模型，即由联合计数 $\tilde{n}$ 上的一组零分布 $M^{n'}$ 构成，分布由

$n' \in B$ $Support\ n \subseteq B$ $n' \in Support\ n$ $P_{M^{n'}}\ \tilde{n} = n' > $ 索引，并

0 $P_{M^{n'}}\ \tilde{n} \in Support\ n = 1$ S^n_{max} n

满足 ( )，且对每个 ( )，( ) 和 (( )) 。另令 为 的一个实值函数。相对于 $\mathcal{M}$ 和 S_{max}， $S(n)$ 的调整指数$AS(n)$ 定义为

$$AS(n) = \frac{S(n) - \mathrm{E}_{M^n}[S(\tilde{n})]}{S^n_{max} - \mathrm{E}_{M^n}[S(\tilde{n})]}.$$

按照约定，当 $S^n_{max} - \mathrm{E}_{M^n}[S(\tilde{n})] = 0$ 时，我们取 $AS(n) = 0$。

注意，模型指标集 B 包含 n的支撑集，以确保零分布 M^n 始终有定义；$M^{n'}$ 对于 n 支撑集中的 n'，其将总概率分配给 n 的支撑集，以确保不可能的联合计数 $\tilde{n}$ 在零假设下不会被赋予正概率；并且零分布 M^n 始终为观测到的 n赋予正概率。

2.2 常见零模型与归一化

三种标准的数据驱动模型分别固定经验边际计数，或基于未合并或合并的经验边际计数固定理论边缘分布，它们是置换模型 $\mathcal{M}_{perm}$ （一种超几何模型）、针对两个总体的独立模型$\mathcal{M}_{2,\ ind}$ （一种多项分布模型），以及针对一个总体的独立模型$\mathcal{M}_{1,\ ind}$ （另一种多项分布模型）。

定义 2.5 ($\mathcal{M}_{perm}$)。 置换模型 $\mathcal{M}_{perm} = \{M^{n'}_{perm}: n' \in \amalg_{I,\ J}\mathbb{N}^{I\times J}\}$ 由如下分布 $M^{n'}_{perm}$ 构成：设 x' 和 y' 是基于 N 个观测值观测到的变量，其联合计数为 n'，边际计数为 u' 和 v'。设 π 是从 N 个观测值中均匀随机选取的一个置换。那么，对于观测值 $\tilde{x}_k = x'_k$ 和$\tilde{y}_k = y'_{\pi(k)}$，联合计数 $\tilde{n}$ 在给定u' 和 v'的条件下服从分布 $M^{n'}_{perm}$。

定义 2.6 ($\mathcal{M}_{2,\ ind}$)。 独立模型 $\mathcal{M}_{2,\ ind} = \{M^{n'}_{2,ind}: n' \in \amalg_{I,\ J}\mathbb{N}^{I\times J}\}$ 由如下方式产生的分布 $M^{n'}_{2\ ,\ ind}$ 构成：令 u' 和 v' 为联合计数 n'的边际计数。给定 u' 和v'，令 $\tilde{x}$ 和 $\tilde{y}$ 为变量，其 N 个观测值分别作为来自概率分布 u'/N，v'/N 的独立简单随机样本获得。那么，给定 u' 和 v'时，$\tilde{x}$ 和 $\tilde{y}$ 的观测值对应的联合计数 $\tilde{n}$ 服从分布 $M^{n'}_{ind}$。

定义 2.7 ($\mathcal{M}_{1,\ ind}$)。一个总体的独立模型 $\mathcal{M}_{1,\ ind} = \{M^{n'}_{1,ind}: n' \in \amalg_{I,\ J}\mathbb{N}^{I\times J}\}$ 由以下方式产生的分布 $M^{n'}_{1\ ,\ ind}$ 构成：令 u' 和 v' 为联合计数 n'的边际计数。在给定 u' 和 v'的情况下，令 $\tilde{x}$ 和 $\tilde{y}$ 为变量，每个变量均有 N 个观测值，且这些观测值分别独立地取自概率分布 $(u' + v') / (2N)$ 的简单随机样本。于是，在给定 u' 和 v'的情况下，$\tilde{x}$ 和 $\tilde{y}$ 的观测值的联合计数 $\tilde{n}$ 服从分布 $M^{n'}_{1\ ,\ ind}$。

S^n_{max} 通常被指定为 $S(n)$ 的上界，尽管这并非严格必要。当S^n_{max} 确实作为最大值获得时，我们将该最大值表示为 $S^n_{max} = \max_{n' \in \mathfrak{X}^n} S(n')$，其中联合计数的某些集合 $\mathfrak{X}^n$ 可能取决于 n，也可能不取决于它。我们考虑 $\mathfrak{X}^n$的几种自然选择，这些选择此前也曾被考虑过 [，见 Cohen，1960年，Warrens， 2013]。

我们可以在指标域上取最大值，此处将其理解为对 N 个观测值所有可能的联合计数

$$\mathfrak{X}_D = \{n' \in \amalg_{I,J} \mathbb{N}^{I \times J} : \sum n'_{ij} = N\}.$$

使用 $\mathfrak{X}^n = \mathfrak{X}_D$ 来定义 *domain maximum* $S^n_{D,\ max}$，只有当观测值的指标在给定观测值数量的情况下达到全局最大值时，才会产生调整后的值 1。请注意，可根据具体情境更新域集的定义，例如，已知边际计数 u 或 v，或者观测值数量 N 未知时。我们还可以考虑模型的支撑集

$$\mathfrak{X}_{M^n} = \{n' \in \mathfrak{X}_D : P_{M^n}(\tilde{n} = n') > 0\}.$$

使用 $\mathfrak{X}^n = \mathfrak{X}_{M^n}$ 来定义 *model maximum* $S^n_{\mathcal{M},\ max}$ 只有当观测值在潜在的数据驱动零分布 M^n的支撑集上为最优时，才会产生调整后的值 1。域最大值和模型最大值在历史上均有应用 [Davenport Jr 和 El-Sanhurry，1991，Cohen，1960] 且已针对指数调整进行了研究 [Warrens， 2013]。

2.3 广义归一化选择的影响 S^n_{max}

通常，若要使调整后的指标 AS 在某个可观测值处取值高达 1，我们需要$S^n_{max} \leq S^n_{D,\ max}$；而要确保调整后的指标 AS 对于根据模型 $\mathcal{M}$先验可观测的联合计数 $\tilde{n}$ 不能取大于 1 的值，则需要 $S^n_{\mathcal{M},\ max} \leq S^n_{max}$。对$S^n_{max}$ 的不同选择会施加不同条件，决定我们将观测到的联合计数 n 视为最优的情形，从而对所得调整后指标 AS赋予不同的解释。下面列举文献中的几个例子：

使用定义域最大值，相对于置换模型对 p 进行调整，可得到 Cohen 的调整指数 κ [Cohen， 1960]，当且仅当变量 x 和 y 完全相同时，该指数取值为 1。使用模型最大值，则可得到 Cohen 所考虑的另一种调整指数 κ/κ_m Cohen [1960],；当观测到的一致性达到边际计数 u 和 v差异所允许的最大程度时，该指数恰好取值为 1。等价地，我们可以在两个总体的独立模型下对 p 进行调整，并使用定义域最大值，从而得到 Cohen 的 κ $\mathcal{M}_{2,\ ind}$。在一个总体的独立模型下对 p进行调整，并使用定义域最大值，可得到 Scott 的 π [Scott、1955、Krippendorff， 1987] $\mathcal{M}_{1,\ ind}$；当变量 x和 y 完全相同时，该指数同样恰好取值为 1，但它假设 x 和 y 来自同一分布，而非如$\mathcal{M}_{2,\ ind}$那样来自不同分布。

在针对置换模型调整 $q(n)$ 时，使用定义域最大值会得到一个缺乏意义的指标：只有当 x 和 y 都将所有观测值分配为同一个值时，该指标才取值 1。相反，标准最大值 $q(n) \leq \frac{1}{2}(q(u) + q(v))$ 与 Rand 指数 $RI = 1 - q\ u\ - q\ v\ + 2q\ n$ [1971], $ARI(\ )(\ )$ [1985] 1 x y ( ) Rand 的定义域最大值相关，可得到标准调整Rand指数 Hubert and Arabie, 。这一调整仅当 和 经过重新标记后完全相同时才取值 。采用非标准最大值 $q(n) \leq \min(q(u)\ q(v))$，Warrens 和 van der Hoef[2022] 指出，这相当于根据成对列联表计算 Loevinger 的 H [Loevinger， 1947, 1948] （见定义 3.1）；由此得到的层次调整 Rand指数仅当 x 和 y 中的一个可以作为另一个变量的函数确定性地获得时，才取值 1。在聚类这一常见情境中，x 和 y 是对同一组观测值进行的两种不同分组所对应的标签，我们认为，只有当这两种分组具有层次结构关系时，层次调整Rand指数才取值 1。据我们所知，文献此前并未明确讨论过针对 q 的模型最大值，而且它在解析上似乎不如其他最大值容易处理。只有当在给定边际计数 u 和 v 的情况下，观测到的成对一致性 $q(n)$ 达到可能的最大值时，该指标才会取值 1。

最后，我们也注意到，这种广义调整概念适用于更广泛的情境：在这些情境中， S^n_{max} 被选定为不是 S 的最大值或上界。例如，传统做法是为所有指标 [固定 $S^n_{max} = 1$ （Albatineh 等人， 2006] ），而 Krippendorff [1987] 则选择 S^n_{max} 作为在给定上下文假设下， n 呈现最大相似性时 S 的取值；该取值不一定与 S的最大值一致。此外，上述框架还涵盖了指标的统计标准化，Romano 等人 [2014, 2016]对此进行了研究。为此，我们取 $S^n_{max} = \mathrm{E}_{M^n}[S(\tilde{n})] + \sigma_{M^n}(S(\tilde{n}))$，其中$\sigma(\cdot)$表示标准差。

3 将一般调整视为算子

为了研究调整算子的有利性质（例如幂等性），接下来我们考察该算子可能以相似方式作用于其上的指数族。从某种意义上说，广义调整算子可理解为具有线性性质，因此线性指标族是值得研究的。经调整后会归并为单一调整指数的线性指数族，最早由 Albatineh 等人 [2006] 针对从成对列联表导出的“成对计数”指标进行研究，随后 Warrens [2008a、b等人 2013], 也对其进行了更一般的研究。我们将以往关于针对非常一般的取值 S^n_{max} 和模型 $\mathcal{M}$进行调整的工作加以拓展。随后，我们将刻画推广可能失效的情境。

3.1 通过线性族获得的良好性质

为将本研究置于基础文献的背景中 [Albatineh 等人， 2006], 我们从成对列联表开始。成对列联表共同评估 x 和 y 中一对观测值被赋予相同或不同 x值，以及分别被赋予相同或不同 y值的频率。一对观测值若被 x 赋予相同的值，且被 y 赋予相同的值，或者被 x 赋予不同的值，且被 y 赋予不同的值，则表明 x 与 y 在如何划分观测值方面具有相似性。若一对观测值被一个变量赋予相同的值，却被另一个变量赋予不同的值，则表明 x 与 y之间存在差异。

定义 3.1 (成对列联表)。根据定义2.1 和2.3，统计一对观测值被赋予 x 和 y 相同或不同取值的成对列联表如下。

	相同 y	不同 y
相同 x	$q(n)\binom{N}{2}$	$(q(u) - q(n))\binom{N}{2}$
不同 x	$(q(v) - q(n))\binom{N}{2}$	$(1 - q(u) - q(v) + q(n))\binom{N}{2}$

用于比较观测值划分的指标（例如通过聚类得到的划分）可能依赖于关联的成对列联表中的联合计数 n，因为它使得能够直接比较那些集合——其标签由 x 和 y 指定，却没有实际意义或不可相互比较。此类指标通常称为成对计数指标 *pair-counting*，与互信息等信息论指数及其变体不同 [Shannon，1948年，Vinh 等人， 2009]。许多成对计数指标采用置换模型进行调整，该模型固定边际计数 u 和 v。注意到成对列联表是 n 的函数，通过 q（n）、 q（u）和 q（v）表示，这就导出了经典的线性指标族，其中 q（n）给定 u 和 v。

定义 3.2 ($\mathcal{L}$; Albatineh 等人 [2006])。 如果对于某些函数 S，则称指标 $\mathcal{L}$属于该族；对于 α 和 β，其定义于 u 和 v 之上时，该指标可改写为

$$S(n) = \alpha(u, v) + \beta(u, v)q(n) \ .$$

这个 f 族 $\mathcal{L}$ 特定于指标 q，也取决于是否固定边际计数 u 和 v。我们定义 e
一类广泛的 具有类似性质的函数族。

定义 3.3 ($\mathcal{L}_{\mathcal{M}}(S)$)。 考虑定义 2.4中的情形。我们定义一个相对于模型 T和指标 $\mathcal{M}$ 的线性指数族 S:

$$\mathcal{L}_{\mathcal{M}}(S) = \left\{T : \forall n \ T(n) = \alpha_n + \beta_n S(n);\ \forall n' \in \mathfrak{X}_{M^n} \ \alpha_n = \alpha_{n'}, \beta_n = \beta_{n'} \neq 0\right\} \ .$$

注意，$\mathcal{L}_{\mathcal{M}_{perm}(q)} = \mathcal{L}$。对于这一一般族，通过要求 α_n 和 β_n 在模型 $\mathcal{M}$中分布 M^n 的支撑集上为常数，我们可以关联指标 $T \in \mathcal{L}_{\mathcal{M}}(S)$ 和 S的期望。在调整过程中，我们也选择通过线性关系关联界限 T^n_{max} 和 S^n_{max}，尽管这样做可能会违背最大值的概念

$$\begin{aligned} \mathrm{E}_{M^n}[T(\tilde{n})] &= \alpha_n + \beta_n \mathrm{E}_{M^n}[S(\tilde{n})] \\ T^n_{max} &= \alpha_n + \beta_n S^n_{max} \ . \end{aligned} \tag{3.1}$$

我们在这一更一般的情形下，对 Warrens [2013] 的研究中的定理 9 和引理 10、12 给出直接的变式。

定理 3.4。 *Given $T \in \mathcal{L}_{\mathcal{M}}(S)$ and S_{max}, define T_{max} according to Equation (3.1). Then the adjustment $AT(n)$ of $T(n)$ with respect to $\mathcal{M}$ and T_{max} is equivalent to the adjustment $AS(n)$ of $S(n)$ with respect to $\mathcal{M}$ and S_{max}: $AT(n) = AS(n)$*。

推论 3.5。 *Consider the context of Theorem 3.4. If S^n_{max} and $E_{M^n}[S(\tilde{n})]$ are each constant in n on the support of each $M^{n'}$,*

then $E_{M^n}[AS\ \tilde{n}\ ] \equiv 0$ $\mathcal{L}_{\mathcal{M}}$ S is closed under A and defining AS_{max} (),() ,,(.)() according to Equation 3.1 A is idempotent ($A^2 = A$) on $\mathcal{L}_{\mathcal{M}}(S)$.

定理 3.4 和推论 3.5 可以重新表述为统计标准化形式：

推论 3.6。 *Let A denote the statistical standardization operator on $\mathcal{L}_{\mathcal{M}}(S)$. That is, if $\tilde{n} \sim M^n$ hasnull distribution given n, A adjusts indices $T \in \mathcal{L}_{\mathcal{M}}(S)$ with respect to $T^n_{max} = E_{M^n}[T(\tilde{n})] + \sigma_{M^n}(T(\tilde{n}))$ and $\mathcal{M}$. Then for all $T = \alpha_n + \beta_n S \in \mathcal{L}_{\mathcal{M}}(S)$ $AT(n) = \text{sgn}(\beta_n)AS(n)$. If $E_{M^n}[S(\tilde{n})]$ and $\sigma_{M^n}(S(\tilde{n})) > 0$ are also each constant in n on the support of each $\mathcal{M}^{n'}$,*

$AT\ \tilde{n}$ has mean 0 and variance 1 with respect ( ) to M^n.

本节所述 $\mathcal{L}_{\mathcal{M}}(S)$ 族的处理方式与 Albatineh 等人对 $\mathcal{L}$ 的原始处理存在显著差异：对于 $T \in \mathcal{L}_{\mathcal{M}}(S)$，我们强制规定，$T^n_{max}$ 是 S^n_{max} 的同一个线性函数，而 T 是 S 的线性函数（公式 3.1）。Albatineh 等人 [2006] 使用 $\mathcal{L}$ 族开展的工作则将 $T^n_{max} = 1$ 固定下来，而不论系数 $\alpha(u, v) = \alpha_n$ 和 $\beta(u, v) = \beta_n$ 取何值；请注意，这种处理要求对 S^n_{max} 有一种完全一般化的理解，因为它甚至可能不是上界。正如该研究及其他研究所指出的，大多数所关注的成对计数指标实际上都将定义域最大值取为 1。然而，为了更一般地考虑对某一指标进行线性变换，这种不自然地固定 $T^n_{max} = 1$ 的做法导致出现由调整后等价指标的系数比 $\frac{1-\alpha}{\beta}$ 参数化的无穷多个子族。它还未能承认通过选择其他最大值可以获得的指标及其解释的丰富性。尽管固定 $T^n_{max} = 1$的约定在文献中很常见 [Warrens, 2008b, Warrens and van der Hoef, 2022, Krippendorff, 1987, Fleiss, 1975], ，但我们采用了可变 T^n_{max} [的约定（Warrens，2013，2008a] ），因为这对于线性变换而言更加自然。

3.2 违反预期性质

C推论3.5 和 3.6 在充分条件下确立了调整的基本期望性质：– 调整后的度量在零假设下均值为 0；– 经过统计标准化的度量在零假设下方差也为 1；– 调整具有幂等性，即根据同一模型和相关最大值进行进一步调整时，调整后的度量不会发生变化（根据公式 3.1）。

充分条件是恒定条件，即对于相对于 $\mathcal{M}$ 和 S_{max} 的调整 A，S^n_{max} 和 $\text{E}_{M^n}[S(\tilde{n})]$ 在每个 $M^{n'}$ 的支撑集上关于 n 都分别保持恒定。这意味着，上界/归一化常数和期望都不能是相对于零分布的随机量。对于域最大值和模型最大值 S_{max} 而言，这一点很直观，并且恒定条件总是得到满足。对于相对于零模型 $\mathcal{M}$ 的期望和标准差而言，由于零分布可能是数据驱动的，这一要求就稍微复杂一些。我们重点讨论期望，标准差的情况与此类似。

$\text{E}_{M^n}[S(\tilde{n})]$在每个 $\mathcal{M}^n$ 的支撑集上为常数这一条件是所有指标均满足的充分条件 S，当分布 $M^n = M$ 不依赖于观测值时 n，或者当分布 M^n 通过某个性质 n 依赖于 $r(n)$ 时，且该性质满足 $r(\tilde{n})\ \ \tilde{n} \sim M^n$ 关于 $\tilde{n}$为常数 $r(\tilde{n}) \equiv r(n)$。第一种情况的一个例子是聚类集成上的均匀分布[Gates 和 Ahn， 2017]。第二种情况要求数据驱动的零分布确定性地固定其纳入的任何观测属性，例如，当边缘分布先验未知时，像置换分布那样固定观测到的边缘分布。

Thus constancy conditions are broadly met by null models which are not data-driven or which depend , only on properties of observed data which are themselves objectively randomness (by the null Desired properties of adjustment follow.

这使得置换模型在历史上的应用显得特殊：置换模型是数据驱动的，但推论 3.5 仍然适用。基于置换模型开展的理论研究未必能顺利推广到新的数据驱动零模型。一般而言，在研究调整后的指数的性质时所考虑的嵌套期望 $\text{E}_{M^n}[\text{E}_{M^{\tilde{n}}}[S(\tilde{n})]]$依赖于两个不同的零分布：$\tilde{n} \sim M^n$ 基于观测到的 n，而 $\tilde{\tilde{n}} \sim M^{\tilde{n}}$ 基于 $\tilde{n}$。这些嵌套期望在恒定条件下自然会化简，从而得到基本的期望性质，但在其他情况下未必如此。特别是，对于数据驱动模型 $\mathcal{M}$ 和指标 S，不应在未经明确验证的情况下假设 $\tilde{n}|n \sim M^n$ 和 $\tilde{\tilde{n}}|\tilde{n} \sim M^{\tilde{n}}$ 满足等式 $\text{E}_{M^n}[S(\tilde{n})] \equiv \text{E}_{M^n}\left[\text{E}_{M^{\tilde{n}}}\left[S(\tilde{\tilde{n}})\right]\right]$。

When a data-driven null model subjects observed distributional properties to randomness desired , properties of adjustment need not hold and must be otherwise confirmed.

事实上，当恒定条件不满足时，我们发现推论3.5的结果会被违反。作为示例证明，我们考虑一个经过传统调整后为非正的指标，以及一个经过统计标准化后确定性地恒等于 0 的指标。

命题 3 7 *there exist .. triples S S_{max} $\mathcal{M}$ such that $E_{M^n}[S\ \tilde{n}\ ]$ is not constant in n on the support of each $\mathcal{M}^{n'}$ and which Consider the context of Theorem 3.4. For each of the following properties, (, ,) () violate the property:*

(1.) for $\tilde{n}|n \sim M^n$ and $\tilde{\tilde{n}}|\tilde{n} \sim M^{\tilde{n}}$, $E_{M^n}[S\ (\tilde{n})\ ] \equiv E_{M^n}\left[E_{M^{\tilde{n}}}[S\ (\tilde{\tilde{n}})\ ]\right]$, (2.) $AS \in \mathcal{L}_{\mathcal{M}}\ (S)$, (3.) $E_{M^n}[AS\ (\tilde{n})\ ] \equiv 0$, and (4.) if S_{max} and AS_{max} are both domain maxima, $A^2S\ (n) \equiv AS\ (n)$ 。

If additionally $S^n_{max} = E_{M^n}[S\ (\tilde{n})\ ] + \sigma_{M^n}\ (S\ (\tilde{n})\ )$ and $\sigma_{M^n}\ (S\ (\tilde{n})\ ) > 0$ for some n, the following property can be violated:

(5.) AS($\tilde{n}$) has mean 0 and variance 1 under the null distribution $\tilde{n} \sim M^n$.

Remark 1: 违反调整下的封闭性（2.）排除了满足推论3.5. 所述幂等性的可能性。具体而言，该推论依据公式（ $AS \in \mathcal{L}_{\mathcal{M}}(S)$ 将 S_{max} 与 AS_{max} 联系起来，并据此定义 A^2。为求完整，我们考察对 S_{max} 和 AS_{max}均使用定义域最大值的情形，并发现修改后的幂等性概念（4.）仍可能不成立。不过，我们注意到，以定义域最大值替代公式（3.1）未必自然，而且可能导致定理3.4失效；正如前文在讨论相对于置换分布和定义域最大值对 $q(n)$ 及 Rand 指数 RI 的调整时所述（第2.3节）。

3.3 违反情形示例

Proof. 命题 3.7 的违反情形。首先，我们给出一个指标 S、定义域最大值 S_{max} 和模型 $\mathcal{M}$ ，使得性质 1-4 均被违反。请注意，性质 1 被违反意味着该命题的非恒定性假设不成立，因此无需单独讨论非恒定性。为提供一个更易处理的示例，我们考虑玩具指标 $S(n) = u_1^2$ 、适用于两个总体的独立模型 $\mathcal{M}_{2,\ ind}$ 以及定义域最大值 N^2。

(1.) 注意，如果 $\tilde{n} \sim M^n_{2,ind}$，那么 $\tilde{u}_1 \sim$ 二项分布$(N, u_1/N)$。由于 $M^n_{2,ind}$ 依赖于 n，嵌套期望无法自然化简；也就是说，$\mathrm{E}_{M^n_{2,ind}}\left[\tilde{u}_1^2\right]$ 在 n 中关于各个 $M_2^{n'}{}_{,ind}$ 的支撑集并非常数：

$$
\begin{aligned}
\mathrm{E}_{M^n_{2,ind}}[\tilde{u}_1] &= N\frac{u_1}{N} = u_1 \\
\sigma^2_{M^n_{2,ind}}(\tilde{u}_1) &= N\frac{u_1}{N}\left(1-\frac{u_1}{N}\right) = u_1\left(1-\frac{u_1}{N}\right) \\
\mathrm{E}_{M^n_{2,ind}}\left[\tilde{u}_1^2\right] &= u_1\left(1-\frac{u_1}{N}\right) + u_1^2 = u_1 + \frac{N-1}{N}u_1^2 \\
\mathrm{E}_{M^n_{2,ind}}\left[\mathrm{E}_{M^{\tilde{n}}_{2,ind}}\left[\tilde{\tilde{u}}_1^2\right]\right] &= \mathrm{E}_{M^n_{2,ind}}\left[\tilde{u}_1 + \frac{N-1}{N}\tilde{u}_1^2\right] \\
&= \frac{2N-1}{N}u_1 + \left(\frac{N-1}{N}\right)^2 u_1^2 \neq \mathrm{E}_{M^n_{2,ind}}\left[\tilde{u}_1^2\right] .
\end{aligned}
$$

(2.) 当 $S_{max} \neq \mathrm{E}^{M^n_{2,ind}}[S(\tilde{n})]$, 时，即当 $u_1 \neq N$ 时，我们计算调整指数：

$$
AS(n) = \frac{u_1^2 - u_1 - \frac{N-1}{N}u_1^2}{N^2 - u_1 - \frac{N-1}{N}u_1^2} = \frac{u_1 - N}{u_1 - N}\cdot\frac{-u_1}{N^2 + (N-1)u_1} .
$$

我们看到 $AS(n)$ 关于 u_1^2 并非线性，具体而言，$AS \notin \mathcal{L}_{\mathcal{M}_{2,ind}}(S)$；因为有理函数 $\frac{-u_1}{N^2 + N-1()u_1}$ ~~无法表示~~为关于常数 α, β 和整数 $0 < u_1 < N$ 的二次函数 $\alpha + \beta u_1^2$ （当 N 较大时）。

(3.) 按照约定，当 $S_{max} = \mathrm{E}^{M^n_{2,ind}}[S(\tilde{n})]$, 时，即当 $u_1 = N$ 时， $AS(n) = 0$ 。因此， $AS(n) \leq 0$ 与 $AS(n) = 0$ 恰好在 $u_1 = 0, N$ 时成立，所以对于 $0 < u_1 < N$，$\mathrm{E}^{M^n_{2,ind}}[AS(\tilde{n})] < 0$ 。

(4.) 根据性质 (3) 推理可得定义域最大值 $AS_{max} = 0$。

调整指数的期望值 $\mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})]$ 以及双重调整指数的值$A^2S(n)$可以计算得出

$$A^2S(n) = \frac{AS(n) - \mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})]}{0 - \mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})]}$$

$$\mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})] = \sum_{k=1}^{N-1} \frac{-k}{N^2 + (N-1)k} \binom{N}{k} \left(\frac{u_1}{N}\right)^k \left(1 - \frac{u_1}{N}\right)^{N-k} .$$

我们可以通过选取 u_1 和 N 并计算（例如），$u_1 = 1$ 和 $N = 2$ 可得 $AS(n) = -\frac{1}{5} \neq -1 = A^2S(n)$。或许更直观的做法是，我们给出整数$1 \leq u_1 < N \leq 100$ 的图，以展示差值通常不为零；参见附录 A中间栏的图 1。

(5.) 对于这一性质，我们考虑另一个玩具指标。令 $S(n) = u_1$, $\mathcal{M} = \mathcal{M}_{2,\ ind}$, $\tilde{n} \sim M_{2,\ ind}^n$ 和 $0 < u_1 < N$。与前面一样，$\tilde{u}_1 \sim$ 服从二项分布 $\mathrm{Binom}(N,\ u_1/N)$，其均值为 u_1，方差为 $u_1(1 - u_1/N) > 0$。对于$0 < u_1 < N$，我们有

$$AS(n) = \frac{u_1 - u_1}{\sqrt{u_1\left(1 - \frac{u_1}{N}\right)}} \equiv 0 .$$

按照约定，即使在 $u_1 = 0,\ N$ 时，$AS(n) = 0$ 也成立。因此，在零分布 $\tilde{n} \sim M_{2,\ ind}^n$ 下，标准化指数 $AS(\tilde{n})$ 恒等于 0，且方差不为 1。 □

Remark 2: 这些关于非零期望（3.）、缺乏幂等性（4.）以及统计标准化失败（5.）的示例证明，表面上似乎从根本上依赖于约定 $AS(n) = 0$ for $S_{max}^n = \mathrm{E}_{M_{2,ind}^n}[S(\tilde{n})]$。之所以在定义 2.4 中采用这一约定，是因为当期望达到可能的最大相似性（$S_{max}^n = \mathrm{E}_{M_{2,ind}^n}[S(\tilde{n})]$）时，它表示除期望（$AS(n) = 0$）之外未观察到任何相似性。如果我们选择不同的约定，$AS(n) = c$ for $S_{max}^n = \mathrm{E}_{M_{2,ind}^n}[S(\tilde{n})]$, 结果仍然成立；参见附录 A。请注意，由于独立模型 $\mathcal{M}_{2,ind}$ 包含某些分布 $M_{2,ind}^n$为定义域中的每个联合计数分配正概率，$\mathfrak{X}_D$ 而另一些分布则将总概率分配给单个联合计数，因此必须选择这样的约定，嵌套期望和双重调整才能得到明确定义。

4 结论

我们研究了在完全一般的模型和归一化常数下指数调整的理论性质，并确立了具备有利性质的充分条件，即一般的指数调整是幂等的，并会得到均值为 0 的调整后指数。

我们还发现，相对于某些模型进行调整，可能会产生违背预期定义性质的调整后指数。当调整中使用的分布由数据驱动时，就会出现这种违背。置换分布，即超几何分布，似乎是一个特例：它们通常虽由数据驱动，却也固定了观测属性，因此满足有利调整性质的充分条件。这使基础性研究 [Albatineh 等人, 2006]聚焦于置换分布的特定性得以凸显，也进一步促使现代 [Gates 和 Ahn, 2017] 以及经典研究 [Krippendorff, 1987] 关注除置换模型之外的模型选择。

未来在这一领域开展的理论和实证研究，将受益于对调整所用模型进行更精确的考量。在可能的情况下，考虑满足推论 3.5 的常数性假设的模型、指数和归一化方式（如 Gates 和 Ahn [2017], 所采用的方式），将有助于确保其具备有利性质。

4.1 未来工作

本研究对指数调整的处理，假设有利性质在零分布下应当严格成立。在实践中，这些性质近似成立或许就足够了。例如，Scott 的 π，即在针对定义域最大值的、含 1 个总体的数据驱动的独立性模型下对 p 进行的调整，不满足恒定条件，且在零分布下的均值不为 0。Scott [1955] 转而给出了大样本刻画。未来的研究或可探究在恒定条件不成立时，有利的渐近行为所需的充分条件。聚类为此类探究提供了一个特别有趣的背景，因为在零分布下，标签的数量可能会变化，而最大可能的

标签可能会随着样本量增加而无限增长。还可以考虑采用不同零模型的模糊聚类 [DeWolfe 和 Andrews, 2025] 也可能值得考虑。

命题 3.7 的第 (3)、(4) 和 (5) 部分允许采用不同的约定 c 来确定何时取最大索引值的期望值。

在命题3.7的证明中，针对性质 (3.)、(4.) 和 (5.) 不成立的情况，我们考虑一种新放宽的约定：当 $\mathrm{E}^{M_{2,ind}^n}[S(\tilde{n})] = S_{max}^n$ 时，将固定值 $AS(n) = c$ 赋给它。作为参考，c 有两个先验看来合乎自然的取值：$c = 0$（原先考虑的取值），这意味着当预期相似性达到最大值时，超出期望的相似度为零；以及 $c = 1$，这意味着当预期相似性达到最大值时，超出期望的相似度为完全相似度，同时确保定义域最大值为 $AS_{max} = 1$。

Proof. **(3.)** 回顾一下 $AS(n) = c$ 对于 $u_1 = N$ 以及对于 $u_1 < N$,

$$AS(n) = \frac{-u_1}{N^2 + (N-1)u_1} .$$

我们计算 AS 在 $0 < u_1 < N$ 时的期望值 $\mathrm{E}^{M_{2,\ ind}^n}[AS(\tilde{n})]$，并注意到对于 $\tilde{n} \sim M_{2,\ ind}^n$, $\tilde{u}_1 \sim$二项分布 $(N,\ u_1/N)$

$$\mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})] = c\left(\frac{u_1}{N}\right)^N + \sum_{k=0}^{N-1} \frac{-k}{N^2 + (N-1)k} \binom{N}{k} \left(\frac{u_1}{N}\right)^k \left(1 - \frac{u_1}{N}\right)^{N-k} .$$

暂且假设 $\mathrm{E}^{M_{2,\ ind}^n}[AS(\tilde{n})] \equiv 0$。那么我们就可以求解 c，其中 $0 < u_1 < N$

$$c = \sum_{k=1}^{N-1} \frac{k}{N^2 + (N-1)k} \binom{N}{k} \left(\frac{N}{u_1} - 1\right)^{N-k} .$$

由于 c 是常数，并非关于 u_1 的严格递减函数，因此不可能有 $\mathrm{E}^{M_{2,\ ind}^n}[AS(\tilde{n})] \equiv 0$。(

4.) 为考察幂等性，我们注意到，对于 $AS(n,\ \max(0, c)$ 是定义域最大值)

$$A^2S(n) = \frac{AS(n) - \mathrm{E}_{M_{ind}^n}[AS(\tilde{n})]}{\max(0, c) - \mathrm{E}_{M_{ind}^n}[AS(\tilde{n})]} .$$

由于 $c = 0$ 在第3.3节中已处理，我们将讨论范围限制为 $c \neq 0$。

令 f、p 和 g 分别表示与 AS 对应的 u_1 的函数（当 $u_1 < N$ 时）、$\tilde{u}_1$ 的二项概率，以及 $AS(\tilde{n})(1 - \delta_{\tilde{u}_1}^N)$ 的期望值，其中 δ_x^y 在 $x = y$ 时为 1，否则为 0。

$$f(u_1, N) = \frac{-u_1}{N^2 + (N-1)u_1} \qquad f_{u_1}(u_1, N) = \frac{-N^2}{(N^2 + (N-1)u_1)^2} < 0$$

$$p(u_1, k, N) = \binom{N}{k} \left(\frac{u_1}{N}\right)^k \left(1 - \frac{u_1}{N}\right)^{N-k} \qquad 1 = \sum_{k=0}^{N} p(u_1, k, N)$$

$$g(u_1, N) = \sum_{k=0}^{N-1} f(k, N) p(u_1, k, N) \qquad \mathrm{E}_{M_{2,ind}^n}[AS(\tilde{n})] = c\left(\frac{u_1}{N}\right)^N + g(u_1, N)$$

Note

$$\frac{-1}{N} < f(N-1, N)(1 - p(u_1, N, N)) \leq g(u_1, N) \leq f(1, N)(1 - p(u_1, N, N)) < \frac{-1}{4N^2},$$

因此，当 $N \to \infty$ 时，$g(u_1, N)$ 一致收敛到 0（$0 < u_1 < N$）。设 $u_1 = N - j$，其中 j 固定。

并注意到 $AS(n) = f(N-j,N) < 0$。然后，用 $\mathbb{1}(\cdot)$ 表示示性函数，我们有 e

$$\begin{aligned}
\frac{1}{N}\frac{A^2S(n)}{AS(n)} &= \frac{1}{N}\left[1 - \frac{c\left(\frac{N-j}{N}\right)^N + g(N-j,N)}{f(N-j,N)}\right] * \\
&\qquad \left(\max(0,c) - c\left(\frac{N-j}{N}\right)^N - g(N-j,N)\right)^{-1} \\
&= \left[\frac{1}{N} + \left(c\left(\frac{N-j}{N}\right)^N + g(N-j,N)\right)\frac{N^2+(N-1)(N-j)}{N(N-j)}\right] * \\
&\qquad \left(\max(0,c) - c\left(\frac{N-j}{N}\right)^N - g(N-j,N)\right)^{-1} \\
&\xrightarrow{N\to\infty} \left[0 + (ce^{-j} - 0)2\right]\left(\max(0,c) - ce^{-j} - 0\right)^{-1} = \frac{2}{e^j\mathbb{1}(c>0) - 1} \neq 0\ .
\end{aligned}$$

我们得出，对于 $N \gg j \geq 1, u_1 = N-j$， $A^2S(n) \approx {}^{-1}_{e^j\mathbb{1}(c>0)-1}$ $AS(n) \approx \frac{-1}{2N}$。因此， $A^2(n) \not\equiv AS(n)$。作为参考，我们提供了整数 $1 \leq u_1 < N \leq 100$ 以及 c 的各种取值下 $AS(n) - A^2S(n)$ 的图（图1）。

(5.) R 回想一下，对于这一性质，我们考虑一个不同的玩具指标 . 令 $S(n) = u_1$ 和 $\mathcal{M} = \mathcal{M}_{2,in}$ $_d$, and recall 对于 $\tilde{n} \sim M^n_{2,ind}$, 我们有 $\tilde{u}_1 \sim$ 二项分布$(N, u_1/N)$。对于 $0 < \tilde{u}_1 < N$, 我们有

$$AS(n) = \frac{\tilde{u}_1 - \tilde{u}_1}{\sqrt{\tilde{u}_1\left(1 - \frac{\tilde{u}_1}{N}\right)}} \equiv 0\ .$$

按照约定， $AS(\tilde{n}) = c$ 当 $\tilde{u}_1 = 0, N$。因此，我们有

$$\begin{aligned}
AS(\tilde{n}) &= c(\delta^0_{\tilde{u}_1} + \delta^N_{\tilde{u}_1}) \\
\mathrm{E}_{M^n_{2,ind}}[AS(\tilde{n})] &= cP_{M^n_{2,ind}}(\tilde{u}_1 = 0, N) = c\left(\left(\frac{u_1}{N}\right)^N + \left(1 - \frac{u_1}{N}\right)^N\right)\ .
\end{aligned}$$

由于对所有 u_1， $\left(\frac{u_1}{N}\right)^N + \left(1 - \frac{u_1}{N}\right)^N > 0$ 都成立，我们有 $\mathrm{E}^{M^n_{2,ind}}[AS(\tilde{n})] \equiv 0$ 当且仅当 $c = 0$。如果 $c = 0$， $AS(\tilde{n})$ 恒等于 0，且方差为 0。 □

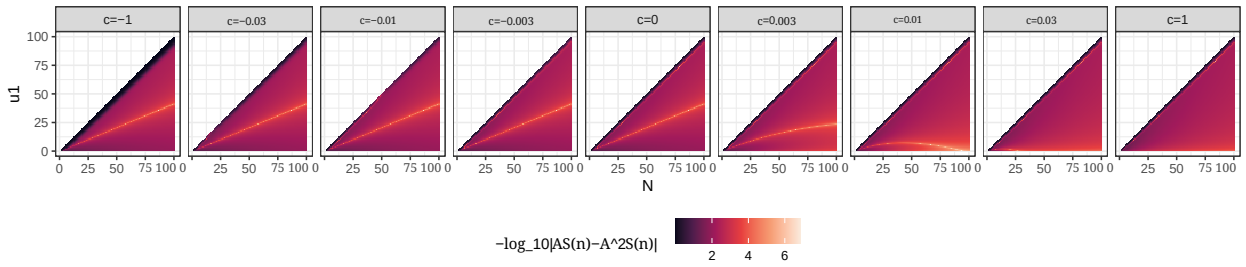


图1：幂等性违背的可视化，即对指标 S 示例而言， AS和 A^2S 普遍不等价的可视化。具体来说，对于 $S(n) = u_1^2$， $S_{max} = N^2$， $AS_{max} = \max(0,c)$，我们展示了 $-\log_{10}|AS(n) - A^2S(n)|$ 关于 u_1、 N 以及当 $S_{max} = \mathrm{E}_{M^n}[S(\tilde{n})]$ 时调整指标 S 数值所采用约定 c 的热图。对数差异越大，表示越接近等价。

致谢

本研究由美国国立卫生研究院通过 R01 HL142049 和 T32 HL007085 资助。

利益冲突

作者声明不存在利益冲突。

参考文献

A. N. Albatineh、M. Niewiadomska-Bugaj 和 D. Mihalko。关于相似性指数及偶然一致性的机会校正。 *J. Classif.*，23(2):301–313，2006年。N. Arinik、V. Labatut 和 R. Figueiredo。用于评估聚类分析和社区检测的外部度量：特征描述与比较。 *IEEE Access*，9:20255–20276，2021年。M. Brusco、J. D. Cradit 和 D. Steinley。71种二元相似系数的比较：基准率的影响。 *Plos one*，16(4):e0247751，2021年。J. Cohen。名义量表的一致性系数。 *Educ. Psychol. Meas.*，20(1):37–46，1960年。E. C. Davenport Jr 和 N. A. El-Sanhurry。Phi/phimax：综述与综合。 *Educ. Psychol. Meas.*，51 4 821 828, 1991 ( ): – 。R. DeWolfe 和 J. L. Andrews。用于调整模糊兰德指数扩展的随机模型。 *Advances in Data Analysis and Classification* 1 25, 2025 ，

页码 – 。L. R. Dice。衡量物种间生态关联程度的方法。

Ecology 26 3 297 302, ，1945

( ): – 。J. L. Fleiss。测量两位评判者对某项特征是否存在的一致性。 *Biometrics*，31 3 651( ): –659，1975年。A. J. Gates 和 Y.-Y. Ahn。随机模型对聚类相似性的影响。

J. Mach. Learn. ，*Res.* 18 87 1 28, 2017

( ): – 。L. A. Goodman 和 W. H. Kruskal。 *Measures of Association for Cross Classifications*，第1章，页码2–34。施普林格统计学系列。施普林格，纽约，NY，1979年。L. Hubert 和 P. Arabie。比较划分。 *J. Classif.*，2:193–218，1985年。L. G. Jeub、O. Sporns 和 S. Fortunato。网络中的多分辨率共识聚类。 *Sci. Rep.*，8 (1):3259，2018年。K. Krippendorff。关联、一致性与公平性。 *Qual. Quant.*，21(2):109–123，1987年。D. Li、C. Ros
´

arXiv e、*preprint arXiv:2407.11371* 2024
A. Yuan 和 C. Zhou。估计序列标注中的偶然一致性。，。J. Loevinger。同质测验的技术与量表分析和因子分析的若干方面之比较。 *Psychol. Bull.*，45(6):507–529，1948年。J. E. Loevinger。构建和评估能力测验的系统方法。

Psychol. ，*Monogr.* 61 4 49, 1947

( ):i– 。M. Meilˇa。比较聚类结果：一种公理化视角。载于 *International Conference on Machine learning*，页码577–584，2005年。W. M. Rand。评估聚类方法的客观标准。 *J. Am. Stat. Assoc.*，66(336): 846–850，1971年。S. Romano、J. Bailey、V. Nguyen 和 K. Verspoor。用于聚类比较的标准化互信息：偶然校正更进一步。载于 *International Conference on Machine Learning* 1143 1151 2014 ，

页码 – 。PMLR，。S. Romano、N. X. Vinh、J. Bailey 和 K. Verspoor。聚类比较度量的偶然校正。 *J. Mach. Learn. Res.*，17(134):1–32，2016年。W. A. Scott。内容分析的信度：名义量表编码案例。 *Public Opin. Q.*，19(3): 321–325，1955年。C. E. Shannon。通信的数学理论。 *Bell Syst. Tech. J.*，27(3):379–423，1948年。

M. Sundqvist、J. Chiquet 和 G. Rigaill。调整后的调整Rand指数：一个多项式故事。*Comput. Stat.*，38(1):327–347，2023年。N. X. Vinh、J. Epps 和 J. Bailey。用于聚类结果比较的信息论度量：机会校正是否必要？载于

International Conference on Machine Learning 1073 1080, ，2009

页码 –。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(4):777–789，2008a。M. J. Warrens。关于 2× 2 表的相似系数及机会校正。 *Psychometrika*，73(3):487–502，2008b。M. J. Warrens。关于关联系数、机会校正和最大值校正。

J. ，*Mod. Math. Front.* 2 4 111

( ): –119，2013年。M. J. Warrens 和 H. van der Hoef。理解调整Rand指数及其他基于对象对计数的划分比较指数。 *J. Classif.*，39(3):487–509，2022年。