arXiv:2304.03380v1 [stat.ME] 6 Apr 2023

边际模型：概述

Tam´as Rudas统计学系，社会科学学院E¨otv¨os Lor´and 大学，布达佩斯trudas@elte.hu

维赫尔 Bergsma 伦敦政治经济学院
W.P.Bergsma@lse.ac.uk

2023年4月10日

摘要

边际模型对一组分类变量的条件关联结构与边际关联结构施加了限制。它们推广了列联表的对数线性模型；对数线性模型是对条件关联结构进行建模的基本工具。本章概述了边际模型在过去20年间的发展。首先介绍一些启发性示例，随后前几节将重点讨论边际模型的定义和特征。具体而言，我们将说明如何根据边际对数线性参数化的性质来理解边际模型的基本性质。接着讨论边际模型的估计算法，重点介绍最大似然和广义估计方程两种方法。我们还将说明边际模型如何帮助理解有向图模型和路径模型，并介绍含有潜变量的边际模型。

1 引言

我们将在第 2 节通过一些示例展开讨论，包括重复测量、缺失数据和图模型。这些示例都涉及这样一类模型的应用：模型仅对变量子集施加限制，也就是说，对包含这些变量联合分布的列联表的边缘分布施加限制。

边际模型施加的限制针对变量子集内的关联结构。关联由在表的边缘分布中计算的对数线性参数来刻画。第 3 节将讨论参数及其参数化的一般方面，包括变异独立性。

边际对数线性参数化在第 4 节中展开介绍，并讨论其一些基本性质，例如变异独立性、光滑性和可塌缩性。根据所选取的边际分布，边际对数线性参数的.....

一旦确定，边际对数线性参数化便适合用于捕捉边际分布与条件关联结构的若干特征。

边际对数线性模型通过将某些边际对数线性参数限制为零来定义，如第 5 节所述。

边际对数线性参数是根据边际分布计算得出的标准对数线性参数，用于衡量条件关联和/或边际关联的强度。边际对数线性参数基于普通优势和优势比，以及它们的高维推广。不过，如第6节所述，其他类型的优势比（对有序数据尤其有用）也可用于定义边际对数线性模型。

第7节包含有关一般类型的条件独立性模型的结果，其中包括某些条件独立性适用于变量子集的情形；这类模型可以表述为边际对数线性模型。

第8节讨论估计与检验。文中介绍并比较了用于最大似然估计的拉格朗日方法和费希尔评分法。文中还介绍了用于估计边际模型的广义估计方程（广义估计方程）方法。

第9节讨论边际对数线性模型的应用领域：这些模型既可为标准分析提供一种通用的实现方式，也可为回答研究问题提供新方法。应用领域包括有向图模型、路径模型和潜变量模型，也提及了许多其他应用。

由于大多数结果均引自研究出版物，因此仅包含极少的证明。

2 动机

有几类统计问题，其中高维联合分布的边际分布发挥着核心作用。本节将讨论其中三类广泛的问题。

2.1 重复测量与面板研究

在许多实验和观察性情境中，会对受试者进行重复测量或重复观察。重复测量的原因包括研究测量的受试者内变异，或通过取平均测量值来降低测量误差。在此类情况下，各次测量的时间间隔较短。重复测量的另一个原因是研究在两次测量之间对受试者实施的处理所产生的效应；在这种情况下，一次测量在处理之前进行，另一次则在处理之后进行。有时，人们关注的是测量结果随时间的变异性或稳定性，而不实施任何处理。

例如，在第一次测量中观测变量 A_1 和 B_1 ，施加处理后，再次测量相同的变量，记为 A_2 和 B_2。有许多相关假设需要检验。第一个假设，例如 H_1，是指在处理前后， A 和 B 都相互独立。可以认为，当且仅当以下两个假设都成立时， H_1 才成立： H_{11} ：“A_1 与 B_1相互独立”，以及 H_{12} ：“A_2 与 B_2相互独立”。这种说法是正确的，但是通常无法通过分别检验假设 H_{11} 和 H_{12}，来构造一个具有给定显著性水平的 H_1 检验。若变量对的样本

A_1， B_1 和 A_2， B_2 相互独立，则可以做到这一点，但在当前的重复测量设置中并非如此。相反，对于样本中的每个个体，都有变量 A_1， B_1， A_2， B_2的观测值； H_1 表明，在这个 4–维分布中存在两个边际独立性，一个涉及 A_1 和 B_1，另一个涉及 A_2 和 B_2。这是一个边际模型。

在此设置中，另一个相关模型，例如 H_2，是指 A 的两次测量的分布相同，即处理不会在总体层面改变分布；对于 B 也是如此，但第二次测量的结果相互独立。因此， H_2 包含了对 $A_1 \times A_2$（边际齐性）、 $B_1 \times B_2$（边际齐性）和 $A_2 \times B_2$ （独立性）边缘分布的限制。

假设 H_1 和 H_2 基于关于联合分布的边际模型。

一种密切相关的纵向设计称为面板研究，参见例如 Frees and Kim（2008年）。在这种研究中，样本中的个体会按照固定间隔接受重复访谈。这种设计的优势[1] 在于，相较于在横断面研究中同时询问当前及以往的立场，它能以更有效的方式研究意见、偏好或态度的变化。这种方法的主要局限在于，人们往往记不清先前的意见或态度，有时也不会如实报告。

在面板数据分析中，从一个立场转变到另一个立场的转移概率尤为重要。尤其是，转移概率是否取决于先前的立场，是一个重要问题，因为这决定了数据的碎片化。更确切地说，假设面板研究有 5 个调查波次，且 A_1， A_2， A_3， A_4， A_5 表示某位受访者在各调查波次对某个特定问题所持的立场，那么人们感兴趣的是判断是否例如，

$$P(A_5|A_4, A_3) = P(A_5|A_4)$$

成立。如果成立，那么除了第 4 波的立场之外，即使把第 3 波的立场也纳入考量，也无法更准确地预测第 5 波的立场。例如，在这种情况下，第 5 波调查时支持某个特定政党的可能性，或许取决于第 4 波调查时偏好的政党；但如果已知后者，那么第 3 波调查时的政党偏好就不会提供额外信息。

稍微一般化一些，若

$$P(A_t|A_{t-1}, A_{t-2}, \ldots, A_1) = P(A_t|A_{t-1})$$

对所有调查波次（时间点） t均成立时，该联合分布称为一步马尔可夫链。很容易看出，这一性质等价于以下条件独立性

$$A_t \perp\!\!\!\perp A_{t-2}, \ldots, A_1 | A_{t-1}.$$

具体而言，对于这 5 个调查波次，这意味着

$$A_3 \perp\!\!\!\perp A_1 | A_2,$$

[1]当然，这种设计也有缺点，包括面板流失，以及即使样本最初选取恰当，随着时间推移，其构成也会不同于当前总体。

$$A_4 \perp\!\!\!\perp A_2, A_1 | A_3,$$

$$A_5 \perp\!\!\!\perp A_3, A_2, A_1 | A_4.$$

对于变量 A_1, A_2, A_3, A_4, A_5 的联合分布，该模型规定在 $A_1 \times A_2$、 $A_1 \times A_2 \times A_3$、 $A_1 \times A_2 \times A_3 \times A_4$ 和 $A_1 \times A_2, \times A_3 \times A_4 \times A_5$ 边际分布上满足条件独立性。

2.2 缺失数据与数据融合

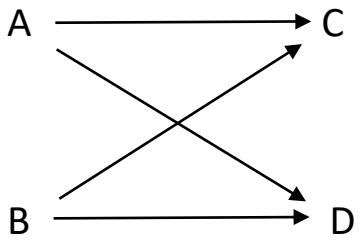
接下来讨论的统计问题引出了这样一项任务：在给定边际分布的情况下生成联合分布。因此，在这些情形中，设计或所收集数据类型所隐含的限制完全确定了某些边际分布（或使其能够被估计），而不仅仅像前面的例子那样为它们指定一个模型。

这类问题中的一类与不完整观测或缺失数据有关（Little 和 Rubin，2019）。通过对人类总体进行调查来收集数据时，抽样程序选中的个体通常并非都会回答问题。有些人无法联系到，有些人虽然联系到了但不愿参与调查，还有些人参与了，却选择不回答某些问题。如何处理完全没有提供任何信息的人是一个棘手的问题，而如何尽可能充分地利用通常只是部分收集到的答案，也是一个重要的统计问题（Little 和 Rubin，2019）。由于参与者退出，实验环境下收集数据时也会出现类似情况。一种方法是考虑针对特定问题子集收集到的回答，并用这些回答来估计答案的联合分布。这些估计是所有答案的联合分布中某些边际分布的估计。采用这一程序是合理的，因为问题子集越小，回答了其中所有问题的个体就越多，其联合分布也就可能比所有变量的联合分布估计得更准确。

例如，设问卷包含 4 道是非题，并设变量A_1、 A_2、 A_3、 A_4 包含这些问题的答案。然后，可以根据对第一道题提供的所有答案来估计 A_1 的分布，其他变量也同样如此。因此，一维边际分布是根据样本的不同子集进行估计的。接下来，根据已经获得的一维边际估计值，以及同时包含 A_1 和A_2回答的观测值，估计 $A_1 \times A_2$ 边际分布。由后者可以估计 A_1与 A_2 之间的优势比（例如参见 Rudas，2018年），并将其与一维边际分布结合起来，以估计 $A_1 \times A_2$ 分布。从理论上讲，可以继续该过程，直到估计出 $A_1 \times A_2 \times A_3 \times A_4$分布，尽管这会引发许多相容性和最优性问题；而且正如稍后将看到的，这一过程是否可行，在很大程度上取决于缺失数据的模式。

有时，缺失数据模式并未被观察到，而是由设计决定的。当问卷过长，或者回答所有问题可能被视为侵犯受访者隐私时，可能会要求一部分受访者回答 A_1， A_2， A_3，另一部分受访者回答 A_1， A_4， A_5；此时，这些可能不是单个问题，而是问题块。当然，也可能采用其他模式。可以估计 $A_1 \times A_2 \times A_3$和 $A_1 \times A_4 \times A_5$边际分布，并据此估计联合分布。这种设计称为拆分问卷

图1：有向无环图



（Rhemtulla 和 Little，2012），但在所谓的基于登记册的人口普查中也会出现类似问题，例如参见 Eppmann 等人（2006）。

在基于登记册的人口普查中（现已有多个国家采用），不再从该国所有居民那里收集信息，而是整合现有登记册（驾驶执照、医疗服务获取等）中的数据，以获取相关信息。各个登记册提供某些条件分布和/或边际分布，任务则是估计联合分布。这个问题称为数据融合（参见 D'Orazio 等，2006），也出现在其他领域，例如 Cocchi（2019）。

2.3 图模型

与有向无环图（也称为贝叶斯网络）相关的图马尔可夫模型广泛应用于专家系统、人工智能，以及一些因果效应建模方法中。

图1展示了一个简单的有向无环图（DAG）。它有四个节点，A，B，C，D，这些节点对应于变量。直观而言，箭头表示直接效应。该图是无环的，因为不存在一串按箭头方向排列、起点和终点相同的节点。

对马尔可夫模型的一种精确定义是：它假设变量之间存在（条件）独立性。正是缺失的箭头意味着由有向马尔可夫性质所定义的条件独立性，参见 Lauritzen（1996）。[2]

与有向无环图相关联的图马尔可夫模型推广了马尔可夫链的条件独立性性质。在马尔可夫链中，条件独立性由缺少时间邻接所蕴含；而在这里讨论的更一般结构中，时间邻接被从图中读取的邻接关系所取代。当图为无向图时，一种称为局部马尔可夫性质的马尔可夫性质指出，给定某个变量的邻居后，该变量与其非邻接节点条件独立。[3] 请注意，这样的

[2]有关因果分析的应用及其与结构方程模型之间的关系，请参见 Bergsma 等人（2009）第 5 章。[3]节点 A 的邻居是指与 A 通过边相连的那些节点。

条件独立性涉及所有变量。

假设变量的联合分布遵循应用于特定图所隐含的局部马尔可夫性质条件独立性，这就是图形对数线性模型，参见 Lauritzen（1996年）。

为定义与有向无环图相关的马尔可夫性质，将有箭头指向 A 的节点称为节点 A 的父节点，并记作 pa(A)。此外，将不存在从 A 出发的有向路径到达它们的节点称为 A 的非后代，并记作 nd(A)。注意，pa(A) $\subseteq$ nd(A)，否则图中就会包含有向环。于是，局部马尔可夫性质为

$$A \perp\!\!\!\perp \mathrm{nd}(A) \setminus \mathrm{pa}(A) | \mathrm{pa}(A).$$

对于图1中的有向无环图，有向马尔可夫性质意味着

$$A \perp\!\!\!\perp B \tag{1}$$

$$C \perp\!\!\!\perp D \,|\, A, B. \tag{2}$$

这两个（条件）独立性中，一个针对 $A \times B$ 边际分布，另一个针对全表。

一般而言，定义与有向无环图相关联的马尔可夫模型的条件独立性，涉及包含某个变量及其非后代的边缘分布。因此，这些模型是边际模型。

根据有向无环图具有马尔可夫性的分布的一个重要性质是，它们可以分解为给定其父节点时各变量的条件分布之积，参见例如Lauritzen（1996年）。对于图1中的有向无环图，其分解为

$$P(ABCD) = P(A)P(B)P(C|A,B)P(D|A,B).$$

3 离散概率分布的参数化

边际模型最便于通过感兴趣变量的联合分布的一种特定参数化来定义，即所谓的边际对数线性参数化，将在第 4 节中讨论。本节将介绍参数和参数化的各种一般特征。

3.1 参数和参数化

参数是分布的任意函数，通常是多维的，即向量值。例如，对于 2×2 分布，使用通常的记法，(p_{11}，p_{12}) 是一个参数，(p_{11}，p_{12}，p_{21}，p_{22}) 或 (p_{11}，p_{1+}，p_{+1}) 也都是参数。统计分析中感兴趣的参数通常表示分布的某种相关特性。

$Outcome$	$Positive$	$Negative$	$Total$
$Treatment$	20	80	100
$Control$	10	90	100

表 1：男性的假设实验结果

例如，表中构成关联的两个变量之间一种广泛使用的关联度量（参见 Rudas（2018））是优势比。

$$\frac{p_{11}p_{22}}{p_{12}p_{21}},$$

它也是分布的一个参数，用于衡量一种无法直接从概率中观察到的特征（关联的强度和方向）。

因此，参数代表了分布中的信息。在许多情况下，人们关注的是能够承载分布中全部信息的参数。更正式地说，可以不考虑某个参数的取值，而是考察生成该参数的函数，并要求该函数可逆。若满足这一条件，该参数就称为参数化。

对于 2×2 分布，(p_{11}，p_{12}) 参数不是参数化，因为即使已知其取值，也无法重构该分布。但 (p_{11}，p_{12}，p_{21}，p_{22}) 和 (p_{11}，p_{1+}，p_{+1}) 参数则是参数化。第一种情况下，单元概率已给出；第二种情况下，则很容易确定。此外，优势比和边际概率 p_{1+} 与 p_{+1} 构成一种参数化，见 Rudas（2018）。在这种情况下，反转参数化，即计算单元概率，需要使用迭代比例拟合或缩放算法等数值算法，参见例如 Rudas（2018）。

参数向量的维度可以任意，但参数化具有最小维度，即分布的维度。对于一个 2×2 分布，虽然有 4 个概率，但这些分布位于四维空间中的一个三维子空间内，因为它们的和为 1。同一事实也可以表述为：在这 4 个概率中，只有 3 个线性独立。因此，参数化的最小维度为 3。

3.2 变异独立性

参数及参数化最理想的性质之一，是其各分量具有变异独立性。在给出一般定义之前，先举一个取自 Rudas 和 Bergsma（2004年）的简单例子来说明这一概念。

假设在一项实验中观察到一张 2×2 处理与结果列联表，并以接受处理者与对照组中阳性结果比例之差作为处理效应的测量指标。假设男性和女性参与者的数据分别如表 1 和表 2 所示。

所选的效应大小度量取值为

$$\frac{20}{100} - \frac{10}{100} = 0.1$$

$Outcome$	$Positive$	$Negative$	$Total$
$Treatment$	60	40	100
$Control$	40	60	100

表2：女性的假设实验结果

针对男性，是

$$\frac{60}{100} - \frac{40}{100} = 0.2$$

对于w 那么，这种处理对女性的效应是否是对男性的两倍?

这个问题的答案并不明确。虽然处理组和对照组中阳性结果概率之差似乎是一个有意义的效应测量指标，而且女性的该指标值是男性的两倍，但也可以作如下推论：男性和女性中，接受处理的个体各有100人，接受对照的个体也各有100人；但男性中有30人、女性中有100人取得阳性结果。鉴于此，效应测量指标的最大可能值为

$$\frac{30}{100} - \frac{0}{100} = 0.3$$

适用于男性并且是

$$\frac{100}{100} - \frac{0}{100} = 1$$

对于女性。因此，男性的治疗效果实际值为其最大可能值的 $1/3$ ，而女性的实际值仅为其理论最大值的 $1/5$ 。

这个例子说明，该度量的可能取值范围会受到其他参数取值的影响。在这种情况下，考虑或不考虑其他参数，对实际取值的评估可能会大不相同。换言之，与其他参数不变异独立的参数缺乏校准。

变异独立意味着上述依赖关系不会发生。如果两个参数的联合取值范围是各自取值范围的笛卡尔积，那么这两个参数就是变异独立的；也就是说，一个参数任何原本可能的取值，都可以与另一个参数任何原本可能的取值相结合。

在这种情 况下，在上文讨论的例子中，治疗效果的度量为

$$\frac{p_{11}}{p_{+1}} - \frac{p_{12}}{p_{+2}} \tag{3}$$

，且它与其他参数 p_{1+} 和 p_{+1}不变异独立。事实上，其最小值为零，最大值为

$$\min\left(1, \frac{p_{1+}}{p_{+1}}\right),$$

因此，其取值范围为 $[0,\min(1, p_{1+}/p_{+1})]$。换言之，度量（3）的取值范围取决于边际分布，而且通常难以确定推断是否应以边缘分布为条件。

优势比与（p_{1+}, p_{+1}）变异独立，因此它是关联参数，其校准不依赖于边缘分布。因此，即使是针对边际分布不同的列联表计算出的优势比，其数值也可以进行比较。男性的优势比为 $1,800/800 = 2.25$，女性为 $3,600/1,600 = 2.25$，这表明处理与阳性结果之间的关联强度在男性和女性中相同；这与对最初计算的度量进行直接比较所得的结论相反。不过，这并不是说优势比作为治疗效果的度量就没有问题，参见 Rudas（2010年）。

对于更高维的列联表，有一些参数化方法依赖于优势比及其推广形式，并具有变异独立性质，由此可自然地定义对数线性模型，参见 Rudas（2018）。不过，对于边际模型的定义，另一种基于边际分布和条件分布的参数化方法更为直接适用。

对于一个 2×2 分布，也可以用 p_{+1} 边际概率以及 $p_{1|1}$ 和 $p_{1|2}$ 条件概率进行参数化。事实上，

$$p_{11} = p_{1|1}p_{+1}$$

$$p_{21} = (1 - p_{1|1})p_{+1}$$

$$p_{12} = p_{1|2}(1 - p_{+1})$$

$$p_{21} = (1 - p_{1|1})(1 - p_{+1}).$$

这种参数化的一个重要特点是，其中的三个参数彼此变异独立。

对一个四维 $A \times B \times C \times D$ 列联表中的分布，也可以采用类似的参数化方法：先对 $A \times B$边际分布进行参数化，再对给定 $A \times B$边际分布时的 $C \times D$条件分布进行参数化。在这里，两组参数彼此变异独立。此外，在这种参数化下，可以施加 A 与 B的边际独立性，然后施加给定 A 和 B时 C与 D的条件独立性。请注意，这恰好就是式（1）和式（2）中定义的边际模型，它由应用于图1所示图的马尔可夫性质所蕴含。

或者，在二元情形下，以下参数构成该四维列联表的一种参数化，其中 OR 表示优势比，COR 表示条件优势比：

$$\theta_1 = (P(A = 1), P(B = 1)$$

$$\theta_2 = \mathrm{OR}(A, B)$$

$$\theta_3 = (P(C = 1|A = 1, B = 1), P(C = 1|A = 1, B = 2),$$
$$P(C = 1|A = 2, B = 1), P(C = 1|A = 2, B = 2), P(D = 1|A = 1, B = 1),$$
$$P(D = 1|A = 1, B = 2), P(D = 1|A = 2, B = 1), P(D = 1|A = 2, B = 2))$$

$$\theta_4 = (\mathrm{COR}(C, D|A = 1, B = 1), \mathrm{COR}(C, D|A = 1, B = 2),$$
$$\mathrm{COR}(C, D|A = 2, B = 1), \mathrm{COR}(C, D|A = 2, B = 2)).$$

在此示例中，θ_1 等同于变量 A 的边际分布以及变量 B 和 θ_3 的边际分布，并给出了在任意可能的类别组合下 C 和 D 的条件分布。

A 和 B。参数 θ_2 是边际分布 $A \times B$ 中的优势比，而 θ_4 是在给定 A 和 B 的所有可能类别组合时，C 和 D 的条件优势比的集合。因此，θ_1 和 θ_2 决定边际分布 $A \times B$，而 θ_3 和 θ_4 决定给定 $A \times B$ 时的条件分布 $C \times D$。

这里，θ_1 和 θ_2 变异独立。此外，θ_3 与 θ_1 和θ_2变异独立，而 θ_4 与其他三个参数均变异独立。同时，θ_2 和 θ_4 与 θ_1 和 θ_3变异独立。

通过固定某个参数化中的部分参数值，可以得到一个统计模型。如果固定的参数与其他参数之间满足变异独立性，那么这不会对其他参数的可能取值产生任何影响；也就是说，对于其他参数的每一种可能取值，模型中都恰有一个分布。换言之，未受限制的参数（分量）可以对限制其他参数（分量）后得到的模型进行参数化。

用一个 2×2、$A \times B$ 列联表的参数化最容易说明这一点，其中$\eta_1 =$(p_{1+}，p_{+1})表示边际分布，而 η_2 表示优势比 OR(A，B)。由于 η_1 和 η_2 变异独立，如果通过施加 $\eta_2 = 1$来定义一个模型，即令 A 和 B独立，那么对于 η_2中边际概率的每一种取值，都恰有一个独立分布。通过将优势比固定为不同的值，可以得到相关模型，参见 Rudas (1991) 和 Rudas 与 Leimer (1992)。

在上述 2×2示例中，当 η_2固定为 1 并得到 2×2 表的独立性模型时，自由度为 1。模型固定了更多参数后，自由度反而更高，这听起来可能有些违反直觉。要接受这一点，就必须记住，自由度与观测频数和期望频数之间的偏差量有关；在决定数据是否提供了反对该模型的证据之前，人们会容忍这种偏差。模型的限制越多，人们在不拒绝模型的情况下愿意容忍的观测频数与期望频数之间的偏差就越大。因此，为了采用具有固定第一类错误概率的检验程序，人们希望使用随模型固定的参数数量单调增加的临界值，而自由度更大的卡方分布具有更大的临界值。因此，与某个模型相关联的自由度并不取决于保留下来自由变化的参数，而是与模型固定的参数数量有关（更准确地说，自由度等于模型固定的参数数量）。

对于这个四维示例，设定 $\theta_2 = 1$意味着 (1)，设定 $\theta_4 =$(1，1、1，1) 则意味着 (2)，由此得到与图1中图相关联的图模型。这两个参数与另外两个参数变异独立，因此，θ_1和 θ_3 参数化了所有满足图1所示图的马尔可夫性质的分布。此外，由于模型固定了 5 个参数值（θ_2 和 θ_4），当模型成立时，标准的皮尔逊统计量和似然比统计量渐近服从自由度为 5 的卡方分布；将数据与最大似然估计进行比较，即可检验模型拟合情况。

按照 Bergsma 和 Rudas（2002a）的定义，本章将边际模型视为上述过程的一种推广。

4 边际对数线性参数化

边际对数线性参数及其参数化推广了上一节讨论的四维列联表示例。边际模型将通过将一些边际对数线性参数设为零来定义。边际对数线性参数和边际对数线性模型的定义，提供了灵活的应用方式，能够捕捉多个分类变量的联合分布的各种有用性质。

4.1 定义

设 $\mathcal{V}$ 是一组分类变量，并令 $\mathcal{M} \subseteq \mathcal{V}$ 表示一个边际分布。下文将根据上下文使用"边际分布"一词，指变量的一个子集、相应列联表的边际表，或由联合分布导出的边际分布。

沿用 Bergsma 和 Rudas（2002a）的定义，边际对数线性参数是对数线性参数（参见 Bishop 等，1975；Agresti，2013年；Rudas（2018）），在表的边际分布中计算得出。为简便起见，本章仅考虑单元概率为正的分布。

变量的每个子集 $\mathcal{E} \subseteq \mathcal{V}$ 都可能对应一个会影响联合分布的效应。为强调这一点，变量的子集也称为效应。效应（与变量子集相关联）的强度可以通过不同方式量化。某种特定的量化方式当然就是一个参数。本节采用一组特定的参数，其他选择将在后文讨论。模型的许多性质并不取决于参数的具体选择；不过，当使用估计的参数值描述模型中的分布时，参数的选择就会产生影响。

效应的经典对数线性参数（例如参见 Bishop 等人（1975）、Agresti（2013年）、Rudas（2018年）） $\mathcal{E}$ 为变量$\mathcal{E}$的每种类别组合 e 指定一个值，记作 $\lambda_e^{\mathcal{E}}$。这些参数通过以下递归定义：

$$\lambda^{\emptyset} = \frac{1}{c_{\mathcal{V}}} \sum_{v} \log P(v),$$

$$\lambda_e^{\mathcal{E}} = \frac{1}{c_{\mathcal{V}\setminus\mathcal{E}}} \sum_{v:(v)_{\mathcal{E}}=e} \log P(v) - \sum_{\mathcal{F}\subsetneq\mathcal{E}} \lambda_{(e)_{\mathcal{F}}}^{\mathcal{F}} \tag{4}$$

其中， e 是变量 $\mathcal{E}$的一个联合类别， $c_{\mathcal{V}\setminus\mathcal{E}}$ 表示 $\mathcal{V}\setminus\mathcal{E}$中变量的联合类别数，而 $(v)_{\mathcal{E}}$表示属于 $\mathcal{E}$中变量的 v的类别。

当所有变量都是二元变量时，可以证明，对数线性参数等于各种平均值；这些平均值是对数的平均值，对应于 $\mathcal{E}$中 l个变量的$l-1$阶条件优势比的根，条件是变量 $\mathcal{V}\setminus\mathcal{E}$的所有可能类别组合（例如，参见Rudas（2018）和本章第6节）。例如，在一个二元 $A\times B\times C$ 表中

$$\lambda_{21}^{AB} = \frac{1}{2}\left(\log P(2,1,1) + \log P(2,1,2)\right)$$

$$
\begin{aligned}
&-\frac{1}{4}\left(\log P(2,1,1)+\log P(2,1,2)+\log P(2,2,1)+\log P(2,2,2)\right)\\
&-\frac{1}{4}\left(\log P(1,1,1)+\log P(1,1,2)+\log P(2,1,1)+\log P(2,1,2)\right)\\
&+\frac{1}{8}\left(\log P(1,1,1)+\log P(1,1,2)+\log P(1,2,1)+\log P(1,2,2)\right)\\
&+\log\left(P(2,1,1)+\log P(2,1,2)+\log P(2,2,1)+\log P(2,2,2)\right)\\
&=\log\sqrt[8]{\frac{P(1,2,1)P(1,2,2)P(2,1,1)P(2,1,2)}{P(1,1,1)P(1,1,2)P(2,2,1)P(2,2,2)}}\\
&=\frac{1}{2}\left(\log\sqrt[4]{\frac{P(1,2,1)P(2,1,1)}{P(1,1,1)P(2,2,1)}}+\log\sqrt[4]{\frac{P(1,2,2)P(2,1,2)}{P(1,1,2)P(2,2,2)}}\right)\\
&=\frac{1}{2}\left(\log\sqrt[-4]{COR(A,B|C=1)}+\sqrt[-4]{COR(A,B|C=2)}\right). \qquad (5)
\end{aligned}
$$

通常，每个效应的对数线性参数都视为向量值参数，其中对于 $\mathcal{E}$ 中变量的每种类别组合，都有一个分量；但为了避免参数各分量之间存在线性依赖，会排除任一变量属于其第一类别的组合。因此，如果 $\mathcal{E}=\{V_1,V_2,\ldots,V_l\}$，且这些变量分别有 c_1，c_2、……、c_l 个类别，那么该对数线性参数具有

$$
(c_1-1)(c_2-1)\cdots(c_l-1) \qquad (6)
$$

分量，并且这些分量通常是线性无关的 .

$\mathcal{E}$ 中变量的第 $l-1$ 阶条件优势比（在以 $\mathcal{V}\setminus\mathcal{E}$ 中变量的任意类别组合为条件时）与 $\mathcal{E}$ 的任意真子集中变量的边际分布变异独立，参见例如 Rudas（2018）。上文通过二元和四元列联表的简单案例对此进行了说明。因此，对数线性参数是条件优势比的函数，常被用作效应中关联程度的度量，该关联无法归因于效应中变量的任何真子集。

在边际分布 $\mathcal{M}$ 中计算 $\mathcal{E}$ 的对数线性参数（$\mathcal{E}\subseteq\mathcal{M}$），意味着使用 $\mathcal{M}$ 的边际概率，而不是 $\mathcal{V}$ 的联合概率。例如，在一个四维二元 $A\times B\times C\times D$ 表中，在 $A\times B\times C$ 边际分布中，AB 效应的边际对数线性参数值为

$$
\frac{1}{2}\sum_{k=1}^{2}\log\left(\frac{P(1,1,k,+)P(2,2,k,+)}{P(1,2,k,+)P(2,1,k,+)}\right)^{1/4}. \qquad (7)
$$

式（7）中的值记为 λ_{AB}^{ABC}，通常记为 $\lambda_{\mathcal{E}}^{\mathcal{M}}$。参数 λ_{AB}^{ABC}是变量 A 和 B之间条件关联的平均值度量（对 C的类别取平均，并对 C的类别取平均），该值是在四维分布的 $A\times B\times C$ 边际分布中计算的。

参数 λ_{AB}^{ABC}只有一个值，因为 A 和 B 都是二元的。比如，如果 B有三个类别，那么该表的大小为 $2\times3\times2\times2$，此时 λ_{AB}^{ABC}有 2 个分量，

一个对应于 A 和 B 的 (2, 2) 指标，另一个对应于 (2,3) 指标。其中，与 (2,2) 相关的优势比由式 (7) 给出，而与 (2, 3) 相关的优势比则取决于所选优势比的类型，参见第6节。这类选择取决于研究问题的特征。

Let

$$\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$$

是一个边际分布序列，使得

$$\mathcal{M}_j \nsubseteq \mathcal{M}_i, \text{ if } i < j$$

and

$$\mathcal{M}_k = \mathcal{V}.$$

这样的序列称为非递减的。在这类边际分布序列中计算的边际对数线性参数在本章中起着核心作用。

为了定义边际对数线性参数，对于每个效应 $\mathcal{E}$，令 $\mathcal{M}(\mathcal{E})$ 表示按非递减顺序排列且包含该效应的第一个边际分布：

$$\mathcal{M}(\mathcal{E}) = \mathcal{M}_i \text{ if } \mathcal{E} \subseteq \mathcal{M}_i \text{ and } \mathcal{E} \nsubseteq \mathcal{M}_j \text{ if } j < i. \tag{8}$$

现在令 $\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})}$ 表示在$\mathcal{M}(\mathcal{E})$ 边际分布中计算的效应 $\mathcal{E}$ 的对数线性参数。这一对数线性参数并非在所有变量的联合分布中计算，而是在边际分布中计算。如上所示，该参数通常是向量值的，但下文将不再特别说明这一点。边际对数线性参数同时衡量边际关联和条件关联的强度。通过选择定义某个效应参数的边际分布，可以忽略一些变量；随后，以边际分布中存在但不属于该效应的变量为条件。关于条件关联与边际关联的进一步讨论，参见 Bergsma 和 Rudas（2013年）。边际对数线性参数 $\{\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})}, \mathcal{E} \subseteq \mathcal{V}\}$ 称为分层且完备的。

此处定义的边际对数线性参数显然包含普通对数线性参数这一特例，也包含多变量逻辑斯蒂参数（McCullagh 和 Nelder，1989年；Glonek 和 McCullagh，1995年），以及 Glonek（1995年）所考察的混合参数。

请注意，上一节末尾定义的参数是边际对数线性参数的一一函数。在本例中，$\mathcal{V} = \{A, B, C, D\}$、$\mathcal{M}_1 = \{A, B\}$ 和 $\mathcal{M}_2 = \{A, B, C, D\}$。因此，$\mathcal{M}(\emptyset) = \mathcal{M}(A) = \mathcal{M}(B) = \mathcal{M}(A, B) = \mathcal{M}_1$ 以及 $\mathcal{M}(C) = \mathcal{M}(A, C) = \mathcal{M}(B, C) = \mathcal{M}(A, B, C) = \mathcal{M}(D) = \mathcal{M}(A, D) = \mathcal{M}(B, D) = \mathcal{M}(A, B, D) = \mathcal{M}(C, D) = \mathcal{M}(A, C, D) = \mathcal{M}(B, C, D) = \mathcal{M}(A, B, C, D) = \mathcal{M}_2$。所指定的参数是对应于这些效应的边际对数线性参数的一一函数。特别地，将 θ_2 和 θ_4 设为 1，等同于设定

$$\lambda_{AB}^{AB} = 0 \text{ and } \lambda_{CD}^{ABCD} = 0.$$

为了获得边际对数线性参数的解析性质，重要的是从中剔除冗余的参数，即那些可以由其他参数计算得出的参数；本章始终假设已进行了这种剔除。公式 (6) 只考虑了参数的非冗余值。

4.2 基本性质

边际对数线性参数具有许多理想性质。对于由 (8) 定义的 $\mathcal{M}$ ：

定理1。 参数 $\{\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})}: \ \mathcal{E} \subseteq \mathcal{V}\}$ 构成变量 $\mathcal{V}$ 的联合分布的一种参数化。

证明。 这是 Bergsma 和 Rudas（2002a）中定理 2 的一部分。严格来说，该证明基于反复应用迭代比例拟合程序，以确定基于指数族混合参数化的联合分布，参见例如 Rudas（2018）。关于混合参数化的一般相关结果由 Barndorff-Nielsen（1978）给出。

□

上述证明中的论证还意味着，对于任意 $1 \leq i \leq k$，在 $\mathcal{M}_1, \ \ldots \mathcal{M}_i$ 中计算得到的边际对数线性参数可用于确定变量 in $\mathcal{M}_i$ 的联合分布。这意味着以下结果。

定理 2。 如果 $\mathcal{M}_i \setminus \cup_{j<i} \mathcal{M}_j \neq \emptyset$，那么边际对数线性参数 $\{\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})}: \ \mathcal{M}(\mathcal{E}) = \mathcal{M}_i\}$ 可确定变量 $\mathcal{M}_i \setminus \cup_{j<i} \mathcal{M}_j$ 的条件联合分布，给定变量 $\mathcal{M}_i \cap (\cup_{j<i} \mathcal{M}_j)$ 的联合分布。

为说明定理 2，对于变量 A、 B、 C，令 $\mathcal{M}_1 = \{AB\}$ 和 $\mathcal{M}_2 = \{ABC\}$。于是 $\mathcal{M}_2 \setminus \mathcal{M}_1 = \{C\}$，而在 $\mathcal{M}_2$ 中计算其边际对数线性参数的效应为 C、 AC、 BC、 ABC，它们对给定联合分布 AB 时 C 的条件分布进行参数化。由于 AB 的联合分布由边际分布 $\mathcal{M}_1$ 参数化，因此两个边际分布中的边际对数线性参数对 ABC 联合分布进行参数化。在两个边际分布中确定的边际对数线性参数是变异独立的。

举一个不那么直观的例子，令 $\mathcal{M}_1 = \{A\}$、 $\mathcal{M}_2 = \{B\}$ 和 $\mathcal{M}_3 = \{ABC\}$。在这种情况下，定理讨论的是给定 AB 时 C 的条件分布，但此时在 $\mathcal{M}_1 = \{A\}$ 和 $\mathcal{M}_2 = \{B\}$ 中确定的边际对数线性参数并不能确定 AB 联合分布，只能确定其一维边际分布。在这种情况下，在 $\mathcal{M}_3$ 中确定的边际对数线性参数中，属于效应 C、 AC、BC、 ABC 的那些参数确定了条件分布。最容易看出这一点的方法是将 AB 边际分布作为第三个边际分布纳入其中；这不会改变这些参数，但会使整个设置与前一个例子基本相同，因为在这种情况下，会先对 AB 联合分布进行参数化，然后再计算 $\{ABC\}$ 边际分布中的参数。

然而，如果 $\mathcal{M}_1 = \{AB\}$、 $\mathcal{M}_2 = \{AC\}$ 和 $\mathcal{M}_3 = \{ABC\}$，那么 $\mathcal{M}_3 \setminus (\mathcal{M}_1 \cup \mathcal{M}_2) = \emptyset$，定理 2 不适用。事实上，如果以 $AB \cup AC$ 为条件，就不再有条件分布剩下。 $\mathcal{M}_3$ 中给出的、对应于效应 BC 和 ABC 的边际对数线性参数，能够确定什么？在这种情况下，它们确定了对 AB 和 AC 边际分布进行参数化所需的额外参数，以便参数化 ABC分布： ABC 的二阶优势比，以及给定 A时 B 和C的条件优势比，参见例如 Rudas（2018）。

	B=1	B=2
A=1	3	1
A=2	1	3

	C=1	C=2
A=1	1	3
A=2	3	1

	C=1	C=2
B=1	3	1
B=2	1	3

表3：边际 al 分布弱相容，但不强相容

一般而言，边际对数线性参数化的所有分量并非都是变异独立的。以下结果给出了层次完备边际对数线性参数化的分量具有变异独立性的必要且充分条件。

定理3。 基于非递减边缘序列的分层且完备的边际对数线性参数化的分量

$$\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k = \mathcal{V}$$

当且仅当以下条件成立时，它们变异独立。要么 $k = 2$ ，要么对于每个 $j = 3$，.....， k，从中取出的极大元素

$$\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_j,$$

say

$$\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_l$$

满足以下条件：要么 $l = 2$ ，要么对于每个 $3 \leq h \leq l$，都存在 $1 \leq g = g(h) \leq h - 1$，使得

$$(\mathcal{H}_1 \cup \ldots \cup \mathcal{H}_{h-1}) \cap \mathcal{H}_h = \mathcal{H}_g \cap \mathcal{H}_h.$$

证明。 这是 Bergsma 和 Rudas（2002a）中的定理4。 □

前一定理中表述的性质称为有序可分解性。如果边缘分布 $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ 关于包含关系两两不可比，也就是说，它们都是极大元素，那么有序可分解性就意味着标准的可分解性概念，参见例如 Rudas（2018）。

例如，在上一处讨论的情形中，对于 $\mathcal{M}_1 = \{AB\}$、 $\mathcal{M}_2 = \{AC\}$ 和 $\mathcal{M}_3 = \{ABC\}$，有序可分解性成立。但对于 $\mathcal{M}_1 = \{AB\}$、 $\mathcal{M}_2 = \{AC\}$、 $\mathcal{M}_3 = \{BC\}$ 和 $\mathcal{M}_4 = \{ABC\}$，有序可分解性不成立，而且很容易找到 $\mathcal{M}_1$ 和 $\mathcal{M}_2$ 中定义的边际对数线性参数的取值，它们会限制 $\mathcal{M}_3$ 中参数的取值范围；参见 Bergsma 和 Rudas（2002a）。表3中列出的三个二元边际（频数）分布是弱相容的，但并非强相容；也就是说，尽管由此生成的一元边缘分布均为均匀分布，却不存在具有这些边缘分布的三元分布。

事实上，如果存在这样的分布，那么对于频数而言，将有 f(1、1、2) ≤ 1(来自 BC 边缘分布) 和 f(1、2、2) ≤ 1 (来自 AB 边缘分布)，但这两个频数之和必须为 3（来自 AC 边缘分布）。这意味着，三个二阶边缘分布，以及相应的边际对数线性参数，并非变异独立。

A=1	C=1	C=2
B=1	t	3-t
B=2	1-t	t

A=2	C=1	C=2
B=1	u	1-u
B=2	3-u	u

表4：具有表3中给出的 AB 和 AC 边缘分布的分布

这是标准对数线性参数与边际对数线性参数之间的一个重要区别。如果在 $\{ABC\}$ 表中计算对数线性参数化，也就是说，采用标准对数线性参数，那么属于 BC 效应的参数本质上是条件优势比 $\mathrm{COR}(B, C|A = a)$，它与 AB 和AC 边际分布变异独立。但如果考虑基于边际分布 $\{AB\}$、$\{AC\}$ 和 $\{BC\}$ 的边际对数线性参数化，那么属于 BC 效应的参数就是边际优势比 $\mathrm{OR}(B, C)$，它与 AB 和AC 边际分布并非变异独立。

要使 AB 和 AC 的边际分布符合表3中的规定，三维列联表必须具有表4所示的结构，这意味着 $t \leq 1$ 和 $u \leq 1$。条件优势比为

$$\mathrm{COR}(B, C|A = 1) = \frac{t^2}{(1-t)(3-t)}$$

and

$$\mathrm{COR}(B, C|A = 2) = \frac{u^2}{(1-u)(3-u)}$$

且其取值不受限制，即根据 t 和 u，可能位于区间 $(0, \infty)$ 内的任意位置。但边际分布优势比是

$$\mathrm{OR}(B, C) = \frac{(t+u)^2}{(4-t-u)^2}$$

且其被限制为不超过 1。

然而，即使在这种情况下，一方面，在边缘分布$\{AB\}$、$\{AC\}$和 $\{BC\}$ 中计算出的边际对数线性参数，与另一方面，在 $\{ABC\}$中计算出的参数，彼此之间也是变异独立的。

4.3 边际对数线性参数的光滑性

边际对数线性模型的定义方式是，假设某些边际对数线性参数为零。这些模型的许多统计学性质，包括最大似然估计的行为以及检验统计量的渐近分布，都取决于所用参数化的解析性质。

如果某个参数作为（概率或频数）分布的函数是连续的、可逆的、二阶连续可微的，且其雅可比矩阵处处满秩，则称该参数是光滑的。

定理4。 分层的完备边际对数线性参数构成频数分布的光滑参数化。

证明。 这是 Bergsma 和 Rudas（2002a）中的定理 2。请注意，只有省略冗余参数值时，光滑性才成立。 □

为了获得概率分布的光滑参数化，必须省略对应于空集的参数 $\lambda_{\emptyset}^{\mathcal{M}(\emptyset)}$，因为根据概率之和必须为 1 这一要求，该参数的值由其他参数决定。

Bergsma 和 Rudas (2002a) 指出（其定理3），对于两个边缘分布 $\mathcal{M}$ and$\mathcal{N}$ 和效应 $\mathcal{E} \subseteq \mathcal{M} \cap \mathcal{N}$，在均匀分布处求值时，参数 $\lambda_{\mathcal{E}}^{\mathcal{N}}$ 和 $\lambda_{\mathcal{E}}^{\mathcal{M}}$ 关于概率分布各分量的偏导数相等，因此这些参数不能同时作为所有分布的光滑参数化的一部分，因为偏导数矩阵并非总是满秩的。Colombi 和 Forcina (2014) 使用一种不同的边际对数线性参数化，对这一问题进行了更详细的分析；该参数化不像式 (7) 那样涉及对条件变量的类别取平均。[4]

下面给出一个结果，用于关联在不同边缘分布中计算的、具有相同效应的边际对数线性参数。

定理 5. 设所有变量均为二元变量，则每个边际对数线性参数都有一个非冗余值。进一步设 $\mathcal{E} \subseteq \mathcal{M} \subset \mathcal{N}$。于是

$$\lambda_{\mathcal{E}}^{\mathcal{N}} = \lambda_{\mathcal{E}}^{\mathcal{M}} + f(\mathbf{\Lambda}_{\mathcal{N}|\mathcal{M}}),$$

对于某个光滑函数 f，其中

$$\mathbf{\Lambda}_{\mathcal{N}|\mathcal{M}} = \{\lambda_{\mathcal{F}}^{\mathcal{N}} : \mathcal{F} \subseteq \mathcal{N}, \mathcal{F} \nsubseteq \mathcal{M}\}.$$

此外，

$$f(\mathbf{\Lambda}_{\mathcal{N}|\mathcal{M}}) = 0 \ \textit{if} \ \big(\mathcal{N} \setminus \mathcal{M}\big) \perp\!\!\!\perp A \,|\, \big(\mathcal{M} \setminus A\big) \tag{9}$$

对于某些 $A \in \mathcal{E}$。

证明。这是 Evans（2015）中定理 3.1 的一部分。 □

例如，该定理的第二个结论表明，如果 $A \perp\!\!\!\perp B|C$，那么 $\lambda_B^{ABC} = \lambda_B^{BC}$。注意到这些对数线性参数是 B 各类别的条件优势的简单函数，即可直接看出这一点。对于第一个参数，条件是 A 和 C；对于第二个参数，条件则只有 C。但是，如果式 (9) 中的条件独立性成立，那么在已对 C 进行条件化的情况下，再对 A 进行条件化并不会提供更多信息，也就是说，

$$P(B = j|C = k) = P(B = j|A = i, C = k),$$

因此，对数线性参数公式中的条件概率是相同的。一般来说，这意味着如果条件（9）对某个分布成立，那么 $\lambda_{\mathcal{E}}^{\mathcal{M}}$和 $\lambda_{\mathcal{E}}^{\mathcal{N}}$不能同时包含在一个光滑参数化中，因为这样雅可比矩阵就不可能是满秩的。

[4]其他参数化方式请参见第 5 节。

4.4 可塌缩性

在给出边际对数线性模型的一般定义之前，我们要讨论的边际对数线性参数化的最后一个性质是可塌缩性。

参数的可塌缩性是一种理想性质，但并非总能实现。可塌缩性有许多变体，它指的是：基于全表进行推断所得的某些方面，与基于相应边际表进行推断所得的结果相同。例如，在一个三维二元表中，$\lambda_{AB}^{ABC}=0$ 通常并不意味着$\lambda_{AB}^{AB}=0$，因此，无论是在全表还是在 AB 边际表中考察，关于变量A 和 B 之间关联强度的推断都不相同。

一般而言，若对于 $\mathcal{M}\subseteq\mathcal{N}$， $\lambda_{\mathcal{E}}^{\mathcal{N}}=\lambda_{\mathcal{E}}^{\mathcal{M}}$成立，则称边际对数线性参数$\lambda_{\mathcal{E}}^{\mathcal{N}}$具有可塌缩性（Ghosh and Vellaisamy，2019）。当然，这一般不可能成立，因为在这种情况下，边际对数线性参数就不会区别于标准对数线性参数。即使对于一种弱得多的要求——方向可塌缩性（即只保留关联方向），Rudas（2015）也证明，存在一种本质上唯一的多元二元分布参数化，对每一种分布都具有方向可塌缩性；它并非对数线性参数化，而是单元概率的线性函数。

因此，可塌缩性通常被理解为一种与参数及特定分布相关，而非单独与参数相关的性质。例如，Ghosh and Vellaisamy（2019）给出了以下结果。

定理6。 设 $\emptyset\neq\mathcal{E}\subseteq\mathcal{M}\subsetneq\mathcal{N}\subseteq\mathcal{V}$ 固定。那么，在二元情形下，若可塌缩性满足

$$\lambda_{\mathcal{F}}^{\mathcal{M}}-\lambda_{\mathcal{F}}^{\mathcal{N}}=0,\ for\ all\ \mathcal{F}\subseteq\mathcal{E}$$

当且仅当对分布 P，

$$\sum_{\mathcal{F}\subseteq\mathcal{E}}\frac{(-1)^{|\mathcal{E}\setminus\mathcal{F}|}}{2^{|\mathcal{M}\setminus\mathcal{F}|}}\sum_{m:(m)_{\mathcal{F}}=(m^*)_{\mathcal{F}}}d(\mathcal{M},m)=0$$

对变量 $\mathcal{M}$ 的所有类别组合 m^* 均成立，其中

$$d(\mathcal{M},m)=\log P_{\mathcal{M}}(m)-\frac{1}{2^{|\mathcal{N}\setminus\mathcal{M}|}}\sum_{n:(n)_{\mathcal{M}}=m}\log P_{\mathcal{N}}(n).$$

证明。 这是 Ghosh and Vellaisamy（2019）中定理 3.1 的一部分。[5] □

5 边际对数线性模型

边际对数线性模型通过对分量应用线性约束，从边际对数线性参数化中获得。如果在第2.1节的示例中，希望假设第一次测量与第二次测量之间的关联强度是

[5]请注意，Ghosh and Vellaisamy（2019）中定理 3.1 的公式 (iii) 似乎存在一个排版错误。

相同，也就是说，处理不影响关联，那么该模型可以通过要求

$$\lambda_{A_1B_1}^{A_1B_1} = \lambda_{A_2B_2}^{A_2B_2}.$$

例如，与图1相关联的图模型（已被反复讨论）等价于第2.3节中(1)和(2)所述的限制。随后，在第3.2节中，考虑了基于这些变量的边缘分布的联合分布参数化，

$$\{AB\},\ \{ABC\},\ \{ABD\},\ \{ABCD\}$$

并证明了定义该模型的限制可以通过限制某些由此得到的参数来施加。在第4.1节中提到，这些限制与

$$\lambda_{AB}^{AB} = 0 \text{ and } \lambda_{CD}^{ABCD} = 0. \tag{10}$$

这就是与图1中有向无环图相关联的图模型的边际对数线性定义。第9.1节将讨论一般意义下图模型的边际对数线性方法。

一般而言，会选取一个非递减边缘序列，并考虑其所隐含的边际对数线性参数化。请记住，参数化中只包含非冗余的参数，因此它是光滑的，参见定理4。在 Bergsma 和 Rudas (2002a) 所考虑的一般情形下，边际对数线性模型通过假设参数属于参数空间的一个线性子空间而得到；当该子空间由令参数等于零的假设所定义时，所得模型就是边际对数线性模型的特殊情形。

这些模型为对数线性模型提供了丰富的推广。模型的实际含义取决于所选的边缘分布以及所施加的限制。本章后文将讨论几个示例。

本节重点介绍边际对数线性模型的一般性质。第一项性质是，这些模型始终存在。

定理 7。 基于边缘分布非递减排序的边际对数线性模型永远不会为空。

*Pr*呃。 这直接由 Bergsma 和 Rudas（2002a 的定理 7 推出). □

一个例子是列联表上的均匀分布，它满足定理中提到的任何边际对数线性模型。请注意，这里不要求变异独立性。

边际对数线性模型所源自的参数化的光滑性（见定理4）意味着，在多项式（参见例如 Rudas（2018））抽样下，通常所期望的渐近性质成立。

定理 8。 假设基于非递减边缘序列的边际对数线性模型包含真实分布。那么，在多项抽样下，随着样本量趋于无穷，真实分布（或其参数）的唯一最大似然估计存在的概率趋于1 。此外，最大似然估计量的渐近分布为正态分布，其期望值等于真实分布。

证明。 这可由 Bergsma 和 Rudas（2002a）的定理 8 得出。 □

这一结果也意味着拟合优度统计量具有标准的渐近行为。

6 边际对数线性模型的替代参数化

优势比有多种定义方式，也可用于对分布进行参数化。这些方式带来了边际对数线性参数化和模型的其他定义，使本章所述方法在解释上更加灵活。

第 4.1 节说明了，边际对数线性参数与局部优势比及其高维推广密切相关（参见例如 Rudas（1998，2018年））。事实上，边际对数线性参数可以由边际表中的局部第 $l-1$ 阶优势比导出。为在边际表中定义这些优势比，设边际 $\mathcal{M}$ 中的边际概率记为 $P_{\mathcal{M}}$，并令 $i_{\mathcal{M}\setminus\mathcal{E}}$ 表示变量 $\mathcal{M}\setminus\mathcal{E}$ 中的一个固定类别。此外，设变量 V_j 的指标为 $1,\ldots,c_j$。则对于边际表中每一个 $(i_1,\ldots,i_l)$：$i_j \geq 2,\ j=1,\ldots,l$，第 $l-1$ 阶局部优势比的形式为

$$\prod_{m_j\in\{0,1\},\,j=1,\ldots,l} P_{\mathcal{M}}^{(-1)^{m_1+\cdots+m_l}}(i_1-m_1,\ldots i_l-m_l,i_{\mathcal{M}\setminus\mathcal{E}}). \tag{11}$$

式（11）是概率或其倒数的乘积。其中涉及的概率位于边际表 $\mathcal{M}$ 中，并且对应于相邻单元格；这些单元格是通过将 $(i_1,\ldots,i_l)$ 中的某些指标减 1 而得到的。式（11）中包含概率还是其倒数，取决于被减小的指标数量的奇偶性。

除了局部优势比（任意阶）之外，也可以使用跨单元格优势比来定义边际对数线性参数。对于一个二元 $I\times J$ 列联表，变量的索引为 0、1、.....、$I-1$ 以及 0、1、.....、$J-1$，跨单元格优势比是 2×2 子表中的优势比，这些子表由参照单元格 (0、0) 和跨越单元格 (i、j) 确定，其中$i=1$、.....、$I-1$，以及 $j=1$、.....、$J-1$。$l-1$ 阶跨单元格优势比的形式为

$$\prod_{m_j\in\{0,i_j-1\},\,j=1,\ldots,l} P_{\mathcal{M}}^{(-1)^{m}}(i_1-m_1,\ldots i_l-m_l,i_{\mathcal{M}\setminus\mathcal{E}}), \tag{12}$$

其中 m 是索引 j 满足 $m_j\neq 0$的数量。在这种情况下，相关单元格通过将某些索引替换为 1 得到。

高阶优势比（无论是局部优势比还是跨单元格优势比）的直观含义，最好通过递归定义来理解，该定义涉及低阶条件优势比的比值。局部优势比衡量相邻单元格之间的关联强度；当变量类别具有顺序时，它们也有意义。跨单元格优势比则衡量将各类别与编码为 1 的参照类别进行比较时的关联强度。

Bartolucci、Colombi 和 Forcina（2007）研究了各种边际交互参数；如果在一个非递减的边缘分布集合中计算这些参数，它们也可以用于定义边际模型。这些广义边际交互是边际概率之和的对数的对比。请注意，本章迄今所讨论的边际对数线性参数，同样是边际概率对数的对比。

Bartolucci 等人（2007）定义交互参数时的核心概念是合并表。局部单元格优势比和跨越单元格优势比通过

从边际表中选取不同的单元格子集，构造二元子表，然后计算这些子集的优势比；而 Bartolucci 等人（2007）的方法则通过合并变量的类别来构造二元子表。合并类别后得到的全局优势比和延续优势比，在变量具有序数性质时尤其有用。由类别已合并的变量构成的表称为合并表。

例如，若考虑一个包含 $I \times J$ 个变量类别的二元边际分布 $\mathcal{M}$，其概率为 $P_{\mathcal{M}}(i,\ j)$，则对每个 $i^* = 2,\ \ldots I$ 和 $j^* = 2,\ \ldots J$，可以考虑以下量：

$$Q_{\mathcal{M},i^*,j^*}(l,l) = \sum_{i=1,\ldots,i^*-1,\, j=1,\ldots,j^*-1} P_{\mathcal{M}}(i,j)$$

$$Q_{\mathcal{M},i^*,j^*}(l,nl) = \sum_{i=1,\ldots,i^*-1,\, j=j^*,\ldots,J} P_{\mathcal{M}}(i,j)$$

$$Q_{\mathcal{M},i^*,j^*}(m,nl) = \sum_{i=i^*,\ldots,I,\, j=1,\ldots,j^*-1} P_{\mathcal{M}}(i,j)$$

$$Q_{\mathcal{M},i^*,j^*}(nl,nl) = \sum_{i=i^*,\ldots,I,\, j=j^*,\ldots,J} P_{\mathcal{M}}(i,j).$$

此处，边际单元概率的求和针对索引小于（l）或不小于（nl）指定的 i^* 和 j^* 的情形进行。

于是，合并表的大小为 2×2，合并后的分布为 $Q_{\mathcal{M},\ i^*,\ j^*}$。这种合并方式将边际表的单元划分为 4 个矩形，并将每个矩形内的概率合并。合并后分布的优势比为

$$\frac{Q_{\mathcal{M},i^*,j^*}(l,l)Q_{\mathcal{M},i^*,j^*}(nl,nl)}{Q_{\mathcal{M},i^*,j^*}(l,nl)Q_{\mathcal{M}}(nl,l)},$$

这称为属于单元（$i^*,\ j^*$）的全局优势比。对于 l 维表，也可以进行类似的合并；所得 2^l 表中的（$l-1$）阶优势比也称为全局优势比。效应 $\mathcal{E}$ 有（c_1-1）（c_2-1）$\cdots$（c_l-1）个全局优势比。

另一种优势比可通过对二元 $I \times J$表进行以下部分合并得到，对每个 $i^* = 1,\ \ldots I-1$和 $j^* = 1,\ \ldots J-1$：

$$R_{\mathcal{M},i^*,j^*}(e,e) = P_{\mathcal{M},i^*,j^*}(i^*,j^*)$$

$$R_{\mathcal{M},i^*,j^*}(n,e) = P_{\mathcal{M}}(i^*+1,j^*)$$

$$R_{\mathcal{M},i^*,j^*}(e,m) = \sum_{j=j^*+1,\ldots,J} P_{\mathcal{M}}(i^*,j)$$

$$R_{\mathcal{M},i^*,j^*}(n,m) = \sum_{j=j^*+1,\ldots,J} P_{\mathcal{M}}(i^*+1,j),$$

其中，e 表示等于，n 表示下一个，而 m 表示大于。

合并后的 2×2 分布所得的优势比，

$$\frac{R_{\mathcal{M},i^*,j^*}(e,e)R_{\mathcal{M},i^*,j^*}(n,m)}{R_{\mathcal{M},i^*,j^*}(e,m)R_{\mathcal{M},i^*,j^*}(n,e)},$$

称为延续优势比。将其写成.....最能看出其含义。

$$\frac{R_{\mathcal{M},i^*,j^*}(n,m)/R_{\mathcal{M},i^*,j^*}(n,e)}{R_{\mathcal{M},i^*,j^*}(e,m)/R_{\mathcal{M},i^*,j^*}(e,e)},$$

即当以第一个变量的下一个类别或当前类别为条件时，第二个变量“延续”而非保持不变的条件优势之比。

在延续优势比的多变量推广中，响应变量会进行合并，但解释变量不会；前提是变量之间存在这种区分。

Bartolucci 等人（2007）将扩展交互参数定义为广义优势比对数的对比，其中包括全局优势比和延续优势比（以及局部单元格优势比和跨越单元格优势比），并指出，对于通过对这些参数施加线性约束而得到的模型，本章迄今介绍的许多结果同样适用。他们将这一更一般的模型类称为层次边际模型。

7 条件独立模型的边际对数线性参数化

许多相关的边际模型假设列联表的各种边缘分布中存在条件独立性。这类模型中最重要的是图模型，其中与有向无环图相关的模型已经讨论过。本章后文将对此作更详细的阐述。本节介绍将条件独立模型表述为边际模型的一般结果，也就是说，通过限制层次完备边际对数线性参数化中的某些参数来构造这类模型。

对于 $i = 1, \ldots\ldots, h$，令 $\mathcal{A}_i \neq \emptyset$、 $\mathcal{B}_i \neq \emptyset$ 和 $\mathcal{C}_i$ 为两两不相交的集合。目标是将以下条件独立性联合表述为边际对数线性模型：

$$\mathcal{A}_i \perp\!\!\!\perp \mathcal{B}_i \,|\, \mathcal{C}_i, \text{ for all } i = 1, \ldots, h. \tag{13}$$

例如，图1中的有向无环图所关联的图模型等价于施加条件独立性(1) 和(2)。在此情况下，$h = 2$, $\mathcal{A}_1 = \{A\}$, $\mathcal{B}_1 = \{B\}$, $\mathcal{C}_1 = \emptyset$ 和 $\mathcal{A}_2 = \{C\}$, $\mathcal{B}_2 = \{D\}$, $\mathcal{C}_2 = \{A, B\}$。

To e探讨何时可将 (13) 表述为边际对数线性模型，定义 e

$$\mathbb{D}_i = \mathbb{P}(\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i) \setminus [\mathbb{P}(\mathcal{A}_i \cup \mathcal{C}_i) \cup \mathbb{P}(\mathcal{B}_i \cup \mathcal{C}_i)],$$

其中 $\mathbb{P}(.)$ 表示幂集。也就是说，对于每个 i， $\mathbb{D}_i$ 是 $\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i$ 的所有子集构成的集合，其中这些子集同时包含来自 $\mathcal{A}_i$ 和 $\mathcal{B}_i$ 的变量。对于有向无环图示例，$\mathbb{D}_2 = \{CD, ACD, BCD, ABCD\}$。以下结果给出了一个充分条件。

定理9。设

$$\mathcal{M}_1, \ldots, \mathcal{M}_k = \mathcal{V}$$

为具有以下性质的非递减边缘序列：

$$\mathcal{C}_i \subseteq \mathcal{M}(\mathcal{E}) \subseteq \mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i, \ \textit{for all } \mathcal{E} \in \cup_{i=1}^{h} \mathbb{D}_i. \tag{14}$$

然后，*(13)* 中的条件独立性定义了一个基于这些边缘分布的边际对数线性模型。更具体地说，分布 P 满足 *(13)* 当且仅当

$$\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})} = 0 \ for \ all \ \mathcal{E} \in \cup_{i=1}^{h} \mathbb{D}_i$$

对于该分布。此外，模型中的分布可由其余边际对数线性参数平滑地参数化：

$$\{\lambda_{\mathcal{E}}^{\mathcal{M}(\mathcal{E})} : \mathcal{E} \notin \cup_{i=1}^{l} \mathbb{D}_i\}.$$

证明。 这是 Rudas 等（2010）中定理1的一部分。 □

条件（14）意味着，对于任何包含来自两个被假定条件独立的变量子集$\mathcal{A}_i$ 和 $\mathcal{B}_i$ 中变量的效应 $\mathcal{E}$，序列中第一个包含 $\mathcal{E}$ 的边际分布必须足够大，能够包含条件集 $\mathcal{C}_i$，但又必须足够小，能够包含于 $\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i$之中。该条件要求边际分布序列足够丰富。

例如，对于由图1中的有向无环图定义的模型，满足性质（14）的边际分布序列有多种选择。显然， AB、 $ABCD$ 序列就是其中一种。但 A、 AB $ABCD$ 或 A、 B、 AB、 ABC、 ABD以及 $ABCD$ 也都是合适的边际分布序列，可以通过将某些边际对数线性参数设为零来定义该模型。然而，定理9并不意味着该有向无环图模型是基于边际分布序列 A、 ACD、ABC、 $ABCD$的边际对数线性模型，因为必须条件独立的变量 C 和 D同时出现，却没有其条件集 AB。对于边际分布序列 A、 AB和 $ABCD$，为指定该有向无环图模型而设为零的效应是

$$AB,\ CD,\ ACD,\ BCD,\ ABCD;$$

第一个在 AB 边际分布中，其余的在 $ABCD$ 边际分布中。请注意，这与式（10）中给出的设定相同；考虑到边际对数线性参数是在边际分布中计算的对数线性参数，因此式（10）中的第二个等式意味着

$$\lambda_{ACD}^{ABCD} = \lambda_{BCD}^{ABCD} = \lambda_{ABCD}^{ABCD} = 0;$$

参见 Rudas（2018年）。

对有向无环图模型中分布进行参数化的边际对数线性参数属于以下效应：

$$\emptyset,\ A,\ B,\ \ C,\ D,\ AC,\ BC,\ \ ABC,\ AD,\ BD,\ ABD.$$

其中，前两个在第一个边际分布中计算，第三个在第二个边际分布中计算，其余的在最后一个边际分布中计算。

定理9的进一步应用将在第9.1节中给出。

定理9中的结果引出了一个问题：给定一组条件独立性，如何确定该模型是否存在光滑边际对数线性定义和参数化？为此，需要考察一系列非递减的

边缘分布，在这些边缘分布中可以表述式（13）所要求的条件独立性，并检验式（14）是否对任意这样的序列都成立。一个显而易见的困难是，某个效应 $\mathcal{E}$ 可能是多个 i 的子集，因此，式（14）可能会对 $\mathcal{M}$（$\mathcal{E}$）施加多项限制。存在光滑边际对数线性定义的一个显而易见的必要条件如下。若对于变量的某个子集 $\mathcal{E}$， $I_{\mathcal{E}}$ 表示满足 $\mathcal{E} \subseteq \mathbb{D}_i$的、从1、……、 h中取出的索引，那么 $\mathcal{M}$（$\mathcal{E}$）应满足

$$\cup_{i \in I_{\mathcal{E}}} \mathcal{C}_i \subseteq \mathcal{M}(\mathcal{E}) \subseteq \cap_{i \in I_{\mathcal{E}}} \mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i$$

并且如果

$$\cup_{i \in I_{\mathcal{E}}} \mathcal{C}_i \nsubseteq \cap_{i \in I_{\mathcal{E}}} \mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i,$$

那么，不存在适当的边际分布序列，且必须满足定理9的充分条件。

Forcina等人（2010）采用不同的方法得出了类似的结果。他们提出了一种算法，用于判断由给定条件独立性集合定义的模型，是否具有此处所讨论意义上的边际对数线性定义，从而是否为光滑模型。[6]

Forcina（2012）进一步探讨了由各种更一般的条件独立陈述集合定义的模型的光滑性问题：对于相关边际分布的任何排列顺序，为了获得指定的条件独立性而必须设为零的边际对数线性参数，无法在效应首次出现的那个边际分布中加以指定。Forcina（2012）讨论的一个例子涉及四个二元变量，并要求

$$X_1 \perp\!\!\!\perp X_2 \,|\, X_3,$$
$$X_2 \perp\!\!\!\perp X_3 \,|\, X_4,$$
$$X_2 \perp\!\!\!\perp X_4 \,|\, X_1.$$

他证明该模型是光滑的，尽管条件（14）不成立，且定理9并不适用。事实上，例如，若这三个边际分布按以下顺序排列，

$$X_1X_2X_3,\ X_2X_3X_4,\ X_1X_2X_4,$$

然后 $X_4 \nsubseteq \mathcal{M}(X_2X_3) = \{X_1X_2X_3\}$。同样，边缘分布的所有其他排序都会导致违反 (14)。

Forcina（2012）提出了一种基于指数族的混合参数化来构造模型中分布的迭代算法，参见例如Rudas（2018），并证明了该算法的收敛性意味着模型的光滑性。

8 估计与检验

在本节中，我们将介绍估计边际对数线性模型的最大似然（ML）方法和广义估计方程（GEE）方法。这两种估计方法都能提供渐近性的……

[6]如果一个模型具有光滑的参数化，则称该模型是光滑的。

无偏估计量，但最大似然估计量相比广义估计方程估计量具有渐近有效的优势。另一方面，广义估计方程方法在计算效率方面更具优势，这一点很重要，因为列联表的规模可能非常大。例如，8 个具有 5 个类别的变量会产生一个大小为 $5^8 = 390{,}625$的列联表。极大似然方法需要估计所有期望单元格频数，而广义估计方程方法只估计观测边际频数的一阶和二阶矩。尽管如此，极大似然方法仍能处理大型表格；例如，我们发现，即使有一百万个单元格的表格，也可以在不太困难的情况下完成估计。

边际对数线性模型的最大似然估计量通常无法用闭式表示，因此需要采用迭代方法。主要有两种方法。首先，有一些算法基于 Aitchison and Silvey（1959）提出的方法，他们采用了拉格朗日乘子技术。Lang and Agresti（1994）首次将此方法用于边际模型，而 Bergsma（1997）提出了一种改进方法，似乎提高了实际应用中的表现。这些方法的一个难点是需要搜索鞍点，因此可能难以监控收敛情况。Bergsma and Rapcsak（2006）通过提出一种替代拉格朗日方法解决了这一问题，该方法将约束最大化问题转化为无约束问题。

极大似然估计的第二种方法是最大化以分层且完备的边际对数线性参数向量表示的似然，例如使用 Fisher评分算法。此方法的缺点是需要“迭代中的迭代”，也就是说，在每一步费希尔评分法中，都需要根据当前估计的边际对数线性参数计算单元概率（这可以通过迭代比例拟合算法完成，该算法保证收敛）。因此，这种方法的计算负担较重，拉格朗日乘子法更具吸引力。为求完整起见，我们仍对这种方法进行说明。

据我们所知，文献中尚未针对一般边际模型描述广义估计方程方法。第8.5节概述了该方法，其中包括对工作协方差矩阵的适当选择。

8.1 边际模型的矩阵表述

设$\mathbf{m}$ 为一个向量，其中包含列联表中各单元格的期望频数。边际对数线性参数 $\boldsymbol{\lambda}$ 可表示为

$$\boldsymbol{\lambda} = \mathbf{B}' \log \mathbf{M}'\mathbf{m} \tag{15}$$

其中$\mathbf{B}$ 和$\mathbf{M}$ 是适当定义的矩阵，撇号表示转置。该表述包含 Bergsma 和 Rudas（2002）的边际对数线性参数化，见第4节。

边际对数线性模型由此定义为

$$\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\beta} \tag{16}$$

对于矩阵$\mathbf{X}$ 和长度小于 $\boldsymbol{\lambda}$ 的参数向量 $\boldsymbol{\beta}$ 。等价地，边际对数线性模型可以指定为

$$\mathbf{C}'\boldsymbol{\lambda} = \mathbf{0} \tag{17}$$

对于适当的矩阵 $\mathbf{C}$。取 $\mathbf{C}$ 为$\mathbf{X}$的正交补集，即$\mathbf{C}'\mathbf{X} = \mathbf{0}$，且 $(\mathbf{X}, \mathbf{C})$ 是一个可逆矩阵，则可知这两种表述是等价的。这些表述被称为自由 和约束 设定（Lang，1996年a）

例如，考虑一个包含期望单元格频数（m_{11}，m_{12}，m_{21}，m_{22}）的 2×2 表。约束指定中的边际齐性模型为 $m_{i+} = m_{+i}$ （$i = 1$，2），其中，下标中的加号表示对该下标求和。用矩阵表示，即为

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \log \left[\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} \right] = 0. \tag{18}$$

在自由度规定中，该模型为 $(m_{i+}, m_{+i}) = (\beta_i, \beta_i)$ $(i = 1,2)$，用矩阵表示为

$$\log \left[\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} \right] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

8.2 ML 估计量的特征化

本节给出一个得分方程和一个拉格朗日得分方程，在某些条件下，它们的解是边际模型的最大似然估计量。求解这些方程的算法将推迟到第8.4节。

令$\mathbf{n}$ 为列联表的观测单元格计数向量。我们假设$\mathbf{n}$ 服从多项式分布或独立泊松分布，其期望频数向量为 $\mathbf{m} = E(\mathbf{n})$。于是，$\mathbf{m}$ 的对数似然为

$$L(\mathbf{m}|\mathbf{n}) = \mathbf{n}' \log(\mathbf{m}) - \mathbf{1}'\mathbf{m} + c \tag{19}$$

其中，$\mathbf{1}$ 是适当长度的全 1 向量，c 是一个常数。在多项分布情形下，约束 $\mathbf{1}'\mathbf{m} = \mathbf{1}'\mathbf{n}$ 成立，但在当前情形下，这不会影响最大似然估计或推断（Lang，1996年）。因此，为了记号简洁起见，下面将忽略多项分布约束。边际对数线性模型下，$\mathbf{m}$ 的最大似然估计量 $\hat{\mathbf{m}}$ 通过在满足形如（16）或（17）的约束条件下最大化对数似然 $L(\mathbf{m}|\mathbf{n})$ 得到。$\mathbf{m}$ 的最大似然估计量 $\hat{\mathbf{m}}$ 已通过两种等价方式得到刻画，即作为以下方程的解：（i）涉及拉格朗日乘子的方程；或（ii）关于 $\boldsymbol{\beta}$ 的得分方程。前一种方法由 Aitchison and Silvey（1958年）和 Lang（1996年）提出，后一种方法则由 Glonek 和 McCullagh（1995年）以及 Colombi 和 Forcina（2001年）研究。

拉格朗日乘子法旨在求出拉格朗日对数似然的驻点

$$L(\mathbf{m}, \boldsymbol{\tau}|\mathbf{n}) = \mathbf{n}' \log(\mathbf{m}) - \mathbf{1}'\mathbf{m} - \boldsymbol{\tau}'\mathbf{C}'\boldsymbol{\lambda}$$

其中，$\boldsymbol{\tau}$ 是拉格朗日乘子向量，$\boldsymbol{\lambda}$ 是形式如（15）的边际对数线性参数。将 $\boldsymbol{\lambda}$ 的雅可比矩阵记为 $\boldsymbol{\Lambda}$，其表达式为

$$\boldsymbol{\Lambda} = \frac{d\boldsymbol{\lambda}'}{d\mathbf{m}} = \mathbf{M}\mathbf{D}_{\mathbf{M}'\mathbf{m}}^{-1}\mathbf{B} \tag{20}$$

其中$\mathbf{D}$ 是其下标位于主对角线上的对角矩阵。对数似然 L 关于$\mathbf{m}$ 求导并令其等于零可得

$$\frac{\mathbf{n}}{\mathbf{m}} - \mathbf{1} + \boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\tau} = \mathbf{0} \tag{21}$$

其中，$\mathbf{n}/\mathbf{m}$ 中的除法是逐元素进行的。

在某些条件下，ML 估计量 $\hat{}\mathbf{m}$ 是方程 (21) 和 (17) 的联立解。充分条件包括：(i) 所有观测频数都严格为正；(ii) 雅可比矩阵 $\boldsymbol{\Lambda}\mathbf{C}$ 具有满列秩。对于大多数（如果不是全部）具有实际意义的边际模型，第二个条件都成立；参见第 4.3 节。然而，在实践中，所有观测频数均为正的条件往往不成立；例如，对于许多现实问题，表格中的单元格数大于样本量，这意味着必定有一些单元格的观测数为零。一种启发式解决方法是将所有观测频数为零的单元格替换为一个很小的常数，使其对似然的总贡献可以忽略不计（Bergsma、Croon和Hagenaars，2009）。

为说明观测单元格为零时的问题，注意，对于由 (18) 定义的边际齐性模型，(21) 变为

$$\frac{n_{ij}}{m_{ij}} - 1 - \frac{\lambda_i}{m_{i+}} + \frac{\lambda_j}{m_{+j}} = 0 \quad i,j = 1,2.$$

C现在考虑 $(i,\ j)$ =(1，1) 的方程。由于 $m_{1+} = m_{+1}$，我们得到 n

$$\frac{n_{11}}{m_{11}} - 1 = 0.$$

解为 $\hat{}m_{11} = n_{11}$，但若 $n_{11} = 0$，则无解。此时真正的极大似然估计量为 $\hat{}m_{11} = 0$；将 n_{11} 替换为一个较小的数，对推断几乎没有影响。

除了拉格朗日乘子法之外，也可以通过 (16) 中关于 $\boldsymbol{\beta}$ 的得分方程来刻画极大似然估计量。似然以 $\boldsymbol{\beta}$ 参数化，通过计算得分方程并求解 $\boldsymbol{\beta}$，即可得到极大似然估计量。如果 $\boldsymbol{\lambda}$ 是边际对数线性参数化，这种方法会更为简便，此时其雅可比矩阵 $\boldsymbol{\Lambda}$ 可逆。对数似然求导后，得到得分向量

$$\mathbf{s}(\boldsymbol{\beta}) := \frac{dL}{d\boldsymbol{\beta}} = \frac{d\boldsymbol{\lambda}'}{d\boldsymbol{\beta}}\frac{d\mathbf{m}}{d\boldsymbol{\lambda}'}\frac{dL}{d\mathbf{m}} = \frac{d\boldsymbol{\lambda}'}{d\boldsymbol{\beta}}\left(\frac{d\boldsymbol{\lambda}'}{d\mathbf{m}}\right)^{-1}\frac{dL}{d\mathbf{m}} = \mathbf{X}'\boldsymbol{\Lambda}^{-1}\Big(\frac{\mathbf{n}}{\mathbf{m}} - \mathbf{1}\Big). \tag{22}$$

$\hat{}\,\hat{}$ 只要所有观测到的单元格频数均为正，极大似然估计量 $\boldsymbol{\beta}$ 就满足 $\mathbf{s}(\boldsymbol{\beta}) = 0$。与拉格朗日乘数法的情形一样，我们建议用一个较小的常数替换观测到的零单元格频数。如果 $\boldsymbol{\lambda}$ 是光滑参数化，那么 $\boldsymbol{\Lambda}$ 可逆，且 $\mathbf{s}(\boldsymbol{\beta})$ $= 0$ 等价于拉格朗日方程（ 21 ），因为 $\mathbf{C}'$ $\mathbf{X}$ $= \mathbf{0}$ 和 $(\mathbf{X},\mathbf{C})$ 是可逆矩阵。计算得分函数可能会耗费大量计算资源，因为需要计算并求逆矩阵 $\boldsymbol{\Lambda}$ 。

8.3 似然比检验与最大似然估计量的渐近分布

假设模型（16）成立。根据 Aitchison and Silvey（1958）和 Lang（1996a）的结果，该模型下的最大似然估计量 $\hat{\mathbf{m}}$ 近似服从大样本多元正态分布，其均值为 $\mathbf{m}$ ，协方差矩阵为

$$\operatorname{cov}(\hat{\mathbf{m}}) \approx \mathbf{D_m} - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{D_m}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}.$$

估计的参数向量 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\log\hat{\mathbf{m}}$ 也具有大样本多元正态分布，均值为 $\boldsymbol{\beta}$ ，协方差矩阵

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\mathbf{m}^{-1}\operatorname{cov}(\hat{\mathbf{m}} - \mathbf{m})\mathbf{D}_\mathbf{m}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

通常的似然比检验可用于选择嵌套模型。设 H_0 和 H_1是嵌套模型，即如果 H_0 为真，则 H_1 为真，并设 $\hat{\mathbf{m}}_k$ 为在 H_k ($k = 0$,1)下对$\mathbf{m}$ 的 ML 估计。对数似然比检验统计量为

$$G^2 = 2\mathbf{n}' \log \frac{\hat{\mathbf{m}}_0}{\hat{\mathbf{m}}_1}$$

在某些正则条件下，如果 H_0 为真，那么 G^2 服从渐近卡方分布，其自由度（df）等于 H_1 的维数减去 H_0的维数。

非嵌套模型可以使用各种信息准则进行比较，例如贝叶斯信息准则（BIC），

$$\mathrm{BIC} = G^2 + 2\,\mathrm{df}\,\log(N)$$

其中 N 为样本量。

8.4 寻找最大似然估计量的算法

8.4.1 拉格朗日方法

已有多种拉格朗日算法被提出，用于求解边际模型的极大似然估计量。在一篇开创性论文中，Aitchison and Silvey（1958）较为一般地介绍了约束最大似然的拉格朗日方法。Lang and Agresti（1994）和 Lang（1996a）将拉格朗日方法引入分类边际模型。本文介绍的算法是对 Bergsma（1997）所开发算法的略微修改；实践经验表明，与原始的 Aitchison and Silvey 算法相比，该算法的收敛性有所改善。

算法的第一步是选择一个适当的起始点 $\mathbf{m}^{(0)}$，之后使用以下公式迭代计算后续估计值 $\mathbf{m}^{(k+1)}$ ($k = 0, 1, 2, \ldots$)

$$\log \mathbf{m}^{(k+1)} = \log \mathbf{m}^{(k)} + step^{(k)}\mathbf{u}\big(\mathbf{m}^{(k)}\big) \tag{23}$$

其中，步长$^{(k)}$ 是适当选择的步长，并且

$$\mathbf{u}(\mathbf{m}) = \frac{\mathbf{n}}{\mathbf{m}} - \mathbf{1} - \mathbf{\Lambda}\mathbf{C}\big(\mathbf{C}'\mathbf{\Lambda}'\mathbf{D_m}\mathbf{\Lambda}\mathbf{C}\big)^{-1}\big[\mathbf{C}'\mathbf{\Lambda}'\mathbf{M}'(\mathbf{n} - \mathbf{m}) + \mathbf{C}'\boldsymbol{\lambda}\big].$$

这里， $\mathbf{\Lambda}$ 由公式（20）定义，并且取决于$\mathbf{m}$。建议的初始值为$\mathbf{m}^{(0)} = \mathbf{n} + \epsilon$，其中 ϵ 是一个较小的常数，例如 10^{-6}。更多详情请参见 Bergsma、Croon和Hagenaars（2009，第2.3.5节）。

Colombi 和 Forcina（2001年）提出了一种密切相关的算法。该算法基于对 (16) 中 $\boldsymbol{\beta}$ 的更新，因此被称为“回归算法”。Evans 和 Forcina（2011年）证明了这两种算法是等价的。他们指出，根据设计矩阵 $\mathbf{X}$ 是否具有由连续协变量产生的分块对角结构，这两种算法的数值特性会有很大差异：如果具有这种结构，回归算法通常效率高得多；如果没有这种结构，Bergsma 算法在实践中通常效率高得多。

尽管我们在实践中观察到算法 (23) 收敛至极大似然估计量的表现相当好，但目前仍缺乏理论结果。一般而言，约束优化问题的收敛性质比无约束优化问题更难确立。拉格朗日乘子法将约束优化问题转化为寻找拉格朗日函数鞍点的问题，但寻找这样的鞍点可能比寻找全局（无约束）最大值或最小值更困难。目前已有两种方法，可将边际模型的极大似然估计问题重新表述为无约束优化问题。

Bergsma 和 Rapcs´ak（2006）提出了一种通用方法，可将带约束的优化问题转化为无约束问题，并将其应用于边际模型的极大似然估计。该算法的优点是具有良好的理论性质，计算效率与式（23）定义的算法相近。

8.4.2 费希尔评分法

本节基于 Colombi 和 Forcina（2001）提出的边际模型 Fisher 评分算法。我们要找出使对数似然（19）最大化的 $\boldsymbol{\beta}$ 在式（16）中的取值。这里， $\boldsymbol{\lambda}$ 是第3节所述的边际对数线性参数化。因此，根据定理4，由式（20）定义的 $\mathbf{\Lambda}$ 可逆。对数似然求导后，得到由式（22）给出的得分向量 $\mathbf{s}(\boldsymbol{\beta})$。关于 $\boldsymbol{\beta}$ 的费希尔信息为

$$\mathbf{I}(\boldsymbol{\beta}) := -E\Big[\frac{d\mathbf{s}(\boldsymbol{\beta})}{d\boldsymbol{\beta}'}\Big] = \mathbf{X}'\mathbf{\Lambda}^{-1}\mathbf{D}_{\mathbf{m}}^{-1}\mathbf{\Lambda}'^{-1}\mathbf{X}.$$

费希尔评分算法如下：

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \mathrm{step}^{(k)}\mathbf{I}(\boldsymbol{\beta}^{(k)})^{-1}\mathbf{s}(\boldsymbol{\beta}^{(k)}). \tag{24}$$

每次迭代时，都需要根据 $\boldsymbol{\lambda}$ 计算期望单元格频数向量 $\mathbf{m}$ ，这可以通过迭代比例拟合算法实现（Bergsma 和 Rudas，2002）。不过，Glonek 和 McCullagh（1995）提出的牛顿-拉夫森迭代方案在数值上可能更高效。

式 (24) 的一个潜在主要数值瓶颈是，每次迭代都需要存储并求逆 $\mathbf{\Lambda}$ 。特别是，如果表中有 K 个单元格，那么 $\mathbf{\Lambda}$ 是一个 $K \times K$ 矩阵。通过直接更新 $\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\beta}$ ，可以得到通常效率高得多的算法。我们

得到更新步骤

$$\begin{aligned}\boldsymbol{\lambda} &\rightarrow \boldsymbol{\lambda} + step\,\mathbf{X}\cdot\mathbf{I}(\boldsymbol{\beta})^{-1}\mathbf{s}(\boldsymbol{\beta})\\ &= \boldsymbol{\lambda} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{D}_{\mathbf{m}}^{-1}\boldsymbol{\Lambda}'^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{D}_{\mathbf{m}}^{-1}(\mathbf{n}-\mathbf{m})\\ &= \boldsymbol{\lambda} - \left[\boldsymbol{\Lambda}'\mathbf{D}_{\mathbf{m}}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}'\mathbf{D}_{\mathbf{m}}\boldsymbol{\Lambda}\mathbf{C}(\mathbf{C}'\boldsymbol{\Lambda}'\mathbf{D}_{\mathbf{m}}\boldsymbol{\Lambda}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Lambda}'\mathbf{D}_{\mathbf{m}}\boldsymbol{\Lambda}\right]\mathbf{D}_{\mathbf{m}}^{-1}(\mathbf{n}-\mathbf{m}),\end{aligned}$$

其中 $\mathbf{C}$ 是$\mathbf{X}$ 的正交补集（见（17））。在实践中，矩阵$\mathbf{C}$ 的列秩通常较低，因此若实现得当，后续的更新步骤会相对高效；详情参见 Colombi 和 Forcina。

总体而言，与拉格朗日算法相比，Fisher评分算法实现起来似乎更加繁琐，尤其是在追求数值效率时。此外，由于需要“迭代中的迭代”，可以预期 Fisher评分算法会比拉格朗日算法更慢。如果采用基于一组非有序可分解边缘分布的参数化，可能会得到超出范围的估计值（负概率）（Colombi 和 Forcina，2001年）。

在更一般的情形下，费希尔评分法的一个缺点是，它要求用感兴趣的参数来表示分布。边际对数线性模型存在此类参数化，但对于 Bergsma（1997）、Lang（2005）以及 Bergsma、Croon和Hagenaars（2009年）所考虑的、基于非对数线性参数的更一般模型，则不存在此类参数化。

8.4.3 软件

以下三个 R 软件包可用于边际建模： `cmm` 由 WicherBergsma 和 Andries van derArk 开发，`mph.fit` 由 JosephLang 开发，以及 `hmmm` 由 Roberto Colombi、Sabrina Giordano 和 ManuelaCazzaro开发。有关 `cmm` 软件包的详细说明请见 `stats.lse.ac.uk/bergsma/cmm/index.html`。该网站包含R 代码，并对Bergsma 等人（2009）中的所有数据示例进行了说明。 `mph.fit` 的文档请见 `homepage.stat.uiowa.edu/~jblang/mph.fitting/index.htm`， `hmmm` 的文档请见 `rdrr.io/cran/hmmm/`；另请参见 Colombi、Giordano和 Cazzaro（2014）。这三个软件包都可以估计各种各样的模型。 `cmm` 的一个特殊功能是能够处理包含潜变量的边际模型，而 `hmmm` 能够处理隐马尔可夫模型和不等式约束。有关 `hmmm`的功能，另请参见第6节。

8.5 广义估计方程方法

边际模型的极大似然估计有一个缺点：需要估计列联表中的所有单元格，因此当单元格数量较大时，计算上难以实现。广义估计方程是一种拟似然方法，它对边际观测之间的协方差矩阵进行建模，同时忽略高阶关联，从而以牺牲部分统计效率为代价提高计算效率。Molenberghs 和 Verbeke（2005，第 8 章）对广义估计方程方法作了详尽的总体概述。大多数关于广义估计方程的文献都使用相关系数对关联进行建模。Lipsitz、Laird 和 Harrington（1991）针对单变量二元响应，提出了基于优势比的广义估计方程方法。Touloumis、Agresti 和 Kateri（2013年）则对多项响应给出了更一般的推导。下文将针对一般边际模型，采用本文所述的广义估计方程程序，即利用对数线性参数对关联进行建模，且感兴趣的

边缘分布可以是多变量的。

广义估计方程方法源自针对多变量边际均值的广义线性模型的得分向量；如果$\mathbf{y} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V}_\mathbf{y}$，并且 $\boldsymbol{\mu} = g(\mathbf{X}\boldsymbol{\beta})$对于某个链接函数 $\boldsymbol{\beta}$成立，则给出 $\boldsymbol{\beta}$ 最大似然估计量的得分方程为

$$\frac{d\boldsymbol{\mu}'}{d\boldsymbol{\beta}}\mathbf{V}_\mathbf{y}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}. \tag{25}$$

如果**y </style}非正态，该方程也可以给出的相合估计量（Wedderburn，1974）。然而，存在一个困难：**$\mathbf{V}_\mathbf{y}$ 通常未知且可能难以估计。Liang和Zeger（1986）提出用可能不正确的˜“工作”协方差矩阵$\mathbf{V}_\mathbf{y}$替代$\mathbf{V}_\mathbf{y}$ ，从而得到广义估计方程

$$\frac{d\boldsymbol{\mu}'}{d\boldsymbol{\beta}}\tilde{\mathbf{V}}_\mathbf{y}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}. \tag{26}$$

˜这里，$\mathbf{V}_\mathbf{y}$ 可以依赖于参数，特别是 $\boldsymbol{\mu}$ 以及描述y相关结构的参数。Liang和Zeger表明，在某些条件下，广义估计方程给出 $\tilde{\boldsymbol{\mu}}$ 的相合估计量 $\boldsymbol{\mu}$。因此，Huber（1967）关于 $\tilde{\boldsymbol{\mu}}$ 协方差矩阵的大样本三明治估计量为

$$\tilde{\text{cov}}(\tilde{\boldsymbol{\beta}}) = \tilde{\mathbf{I}}^{-1}\tilde{\mathbf{J}}\tilde{\mathbf{I}}^{-1}$$

其中

$$\tilde{\mathbf{I}} = \frac{d\boldsymbol{\mu}'}{d\boldsymbol{\beta}}\tilde{\mathbf{V}}_\mathbf{y}^{-1}\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} \qquad \tilde{\mathbf{J}} = \frac{d\boldsymbol{\mu}'}{d\boldsymbol{\beta}}\tilde{\mathbf{V}}_\mathbf{y}^{-1}\mathbf{V}_\mathbf{y}^{*}\tilde{\mathbf{V}}_\mathbf{y}^{-1}\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}.$$

这里，$\mathbf{V}_\mathbf{y}^{*}$ 是$\mathbf{V}_\mathbf{y}$的相合估计量。

现在给出用于估计边际模型 (16) 中 $\boldsymbol{\beta}$ 的广义估计方程方法，其中边际观测频数向量记为$\mathbf{y} = \mathbf{M}'\mathbf{n}$ ，相应的期望频数向量记为 $\boldsymbol{\mu} = E(\mathbf{y}) = \mathbf{M}'\mathbf{m}$。于是

$$\mathbf{V}_\mathbf{y} = \mathbf{M}'\mathbf{D}_\mathbf{m}\mathbf{M} - N^{-1}\boldsymbol{\mu}\boldsymbol{\mu}' \tag{27}$$

其中 $N = \mathbf{1}'\mathbf{n}$ 是样本量。我们可以将边际模型 (16)（其中 $\boldsymbol{\lambda}$ 由 (15) 给出）写作

$$\boldsymbol{\mu} = \exp(\mathbf{U}\mathbf{X}\boldsymbol{\beta})$$

其中$\mathbf{U}$ 是$\mathbf{B}$ 的正交补集，也就是说，$\mathbf{B}'\mathbf{U} = \mathbf{0}$ ，且 $(\mathbf{B}, \mathbf{U})$ 是一个可逆矩阵。因此，

$$\frac{d\boldsymbol{\mu}'}{d\boldsymbol{\beta}} = \mathbf{X}'\mathbf{U}'\mathbf{D}_{\boldsymbol{\mu}}$$

因此 (25) 变为

$$\mathbf{X}'\mathbf{U}'\mathbf{D}_{\boldsymbol{\mu}}\tilde{\mathbf{V}}_\mathbf{y}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}. \tag{28}$$

一个难点是，$\mathbf{V}_{\mathbf{y}}$ 通常不可逆，在这种情况下，我们可以用以下公式替代 (28)：

$$\mathbf{y} - \boldsymbol{\mu} + \mathbf{V}_{\mathbf{y}}\mathbf{D}_{\boldsymbol{\mu}}^{-1}\mathbf{B}\mathbf{C}\boldsymbol{\tau} = \mathbf{0} \tag{29}$$

其中，$\boldsymbol{\tau}$ 是一个待估参数。简单计算可知，若$\mathbf{V}_{\mathbf{y}}$ 可逆，则式（28）和式（29）等价。注意，式（29）可由拉格朗日得分方程（21）的左右两边分别左乘$\mathbf{M}'\mathbf{D}_{\tilde{\mathbf{m}}}\mathbf{V}_{\mathbf{y}}^{-1}$得到。将式（29）中的$\mathbf{V}$替换为工作协方差矩阵$\mathbf{V}_{\mathbf{y}}$，即可得到用于边际模型的广义估计方程。计算$\mathbf{J}$需要$\mathbf{V}_{\mathbf{y}}$ 的相合估计量$\mathbf{V}_{\mathbf{y}}^{*}$，为此可以取$\mathbf{V}_{\mathbf{y}}^{*} = \mathbf{M}'\mathbf{D}_{\mathbf{n}}\mathbf{M}$。

接下来需要确定一个工作协方差$\mathbf{V}_{\mathbf{y}}$。一种简单的方法如下。假设边际模型基于非嵌套边际 $\mathcal{M}_1, \ldots, \mathcal{M}_k$。那么，由式（27）给出的$\mathbf{V}_{\mathbf{y}}$ 是以下边际的期望边际频数的函数

$$\{\mathcal{M}_i \cup \mathcal{M}_j | i, j = 1, \ldots, k\}. \tag{30}$$

一种简单的工作协方差矩阵可通过假设边际分布 $\mathcal{M}_i \cup \mathcal{M}_j$服从一个（可能不正确的）条件独立模型来得到：

$$(\mathcal{M}_i \setminus \mathcal{M}_j) \perp\!\!\!\perp (\mathcal{M}_j \setminus \mathcal{M}_i) | \mathcal{M}_i \cap \mathcal{M}_j.$$

这给出了 $\mathcal{M}_i \cup \mathcal{M}_j$ 中期望边际频数关于 $\mathcal{M}_i$ 中期望边际频数的闭式表达式，因此可以利用例如牛顿–拉弗森方法求解满足 (16) 的 (29)，得到 $\boldsymbol{\beta}$。

8.5.1 关于广义估计方程方法的说明

如果工作协方差矩阵不正确，与渐近最优的极大似然方法相比，广义估计方程方法会损失渐近效率。上文提出了一种简单的工作协方差；对于单变量边际分布，这相当于采用独立性工作假设。Touloumis 等人（2013年）指出，如果边际观测之间存在较强依赖，这种做法可能会造成较大的效率损失。通过指定更接近真实情况的工作协方差矩阵，可以提高效率；具体做法是为 (30) 中的边际分布指定并估计一个合适的参数模型。Touloumis 等人通过使用齐次关联模型对二元边际进行建模（关联模型概述见 Agresti，2013年，第9章，或 Forcina 和 Kateri，2021年），显著提高了单变量边际模型的效率。

广义估计方程方法是一种拟似然 方法。另一种广受欢迎的拟似然方法是复合似然，它基于将某些边际分布似然相乘而定义的拟似然；有关概述，请参见 Molenberghs 和 Verbeke（2005 年，第 9 章）。复合似然的优势在于，它既可用于边际分布模型，也可用于条件模型。广义估计方程方法的优势在于，通过改进工作协方差矩阵的设定，其渐近效率可以任意接近极大似然方法的效率。

使用广义估计方程估计进行模型比较比使用极大似然估计更加困难。Rotnitzky 和 Jewell（1990年）开发了模型比较和拟合优度工具，而 Pan（2001年）开发的拟似然信息准则（QIC）尤其受欢迎。

9 应用领域

9.1 有向图模型

与有向无环图或更一般的链图（Lauritzen，1996年）相关的分类数据图模型，在 Bergsma 和 Rudas（2002）的意义上属于边际对数线性模型。近年来，这些模型的参数化受到广泛关注，参见 Rudas、Bergsma 和 N´emeth（2010年）、Marchetti 和 Lupparelli（2011年）、Evans 和 Richard- son（2013年）、N´emeth 和 Rudas（2013年），以及 Nicolussi 和 Colombi（2017年）。对于有向无环图，马尔可夫性质是

$$V_i \perp\!\!\!\perp \mathrm{nd}(V_i) \mid \mathrm{pa}(V_i). \tag{31}$$

这里，对每个变量 V_i，nd(V_i)表示非后代，pa(V_i)表示 V_i的父节点。Rudas 等（2010）给出的这类模型的边际对数线性参数化，基于变量的良序编号（Lauritzen 等，1990），使得（31）等价于

$$V_i \perp\!\!\!\perp \mathrm{pre}(V_i) \setminus \mathrm{pa}(V_i) \mid \mathrm{pa}(V_i), \tag{32}$$

其中，pre(V_i)是良序编号中位于 V_i 之前的变量集合。Rudas 等（2010）提出的参数化基于边缘分布 $\{V_i\} \cup$pre(V_i)，从而可以按照定理1进行参数化。

关于链图相关统计模型的早期研究包括 Lauritzen 和 Wermuth（1989年）、Frydenberg（1990年）、Cox 和 Wermuth（1996年）、Andersson、Madigan 和 Perlman（2001年）、Richardson（2003年）、Wermuth 和 Cox（2004年），以及 Drton（2009）。对于链图的一个分量 $\mathcal{K} \subseteq \mathcal{V}$，ND($\mathcal{K}$)是 $\mathcal{K}$的非后代集合，即满足以下条件的那些分量的并集：不存在半有向路径从 $\mathcal{K}$中的任何节点通往这些分量中的任何节点；其中不包括 $\mathcal{K}$。PA($\mathcal{K}$)是 $\mathcal{K}$的父节点集合，即有箭头指向 $\mathcal{K}$中某个节点的那些分量的并集。 $\mathcal{X} \subseteq \mathcal{K}$的邻居集合nb($\mathcal{X}$)，是 $\mathcal{K}$中与 $\mathcal{X}$中的某个节点相连的节点集合；pa($\mathcal{X}$)则是有箭头指向 $\mathcal{X}$中任一节点的节点集合。

链图模型由以下某些性质的组合定义。

P1 对所有分量 $\mathcal{K}$, $\mathcal{K} \perp\!\!\!\perp \{\mathrm{ND}(\mathcal{K}) \setminus \mathrm{PA}(\mathcal{K})\} \mid \mathrm{PA}(\mathcal{K})$,

P2a 对所有 $\mathcal{K}$ 和 $\mathcal{X} \subseteq \mathcal{K}$, $\mathcal{X} \perp\!\!\!\perp \{\mathcal{K} \setminus \mathcal{X} \setminus \mathrm{nb}(\mathcal{X})\} \mid \{\mathrm{PA}(\mathcal{K}) \cup \mathrm{nb}(\mathcal{X})\}$,

P2b 对所有 $\mathcal{K}$ 和 $\mathcal{X} \subseteq \mathcal{K}$, $\mathcal{X} \perp\!\!\!\perp \{\mathcal{K} \setminus \mathcal{X} \setminus \mathrm{nb}(\mathcal{X})\} \mid \mathrm{PA}(\mathcal{K})$,

P3a 对所有 $\mathcal{K}$ 和 $\mathcal{X} \subseteq \mathcal{K}$, $\mathcal{X} \perp\!\!\!\perp \{\mathrm{PA}(\mathcal{K}) \setminus \mathrm{pa}(\mathcal{X})\} \mid \{\mathrm{pa}(\mathcal{X}) \cup \mathrm{nb}(\mathcal{X})\}$,

P3b 对所有 $\mathcal{K}$ 和 $\mathcal{X} \subseteq \mathcal{K}$, $\mathcal{X} \perp\!\!\!\perp \{\mathrm{PA}(\mathcal{K}) \setminus \mathrm{pa}(\mathcal{X})\} \mid \mathrm{pa}(\mathcal{X})$.

I型马尔可夫性质（P1、P2a、P3a）也称为Lauritzen–Wermuth– Fryden- berg块递归马尔可夫性质，参见LauritzenandWermuth（1989年）和Fryden-berg（1990年）；II型马尔可夫性质（P1、P2a、P3b）也称为Andersson– Madigan–Perlman块递归马尔可夫性质，参见Andersson等（2001年）。

I 型模型的光滑性可由 Frydenberg（1990年）的结果推出，也可通过应用定理 1 轻松得到。

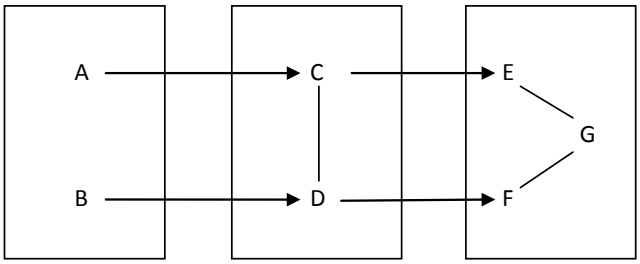


图2：根据 Andersson–Madigan–Perlman 的解释，该链图对应一个光滑模型

以下示例说明，边际对数线性参数化可用于证明属于同时包含非光滑模型的模型类的链图模型的光滑性。图2中采用类型II解释的图是一个光滑模型，并且可以使用边缘分布 AB, ABC, ABD, CDE, CDF, CDG, $CDEG$,$CDFG$, $CDEFG$, $ABCDEFG$ 进行参数化。一般而言，类型II模型并不光滑，见Drton (2009)，但在本例中，定理1立即表明其具有光滑性。

Drton（2009）证明，IV型模型（P1、P2b、P3b）是光滑的，并给出了参数化。Lupparelli、Marchetti 和 Bergsma（2009）通过示例说明，这些模型是边际对数线性模型。现在，我们将定理1中的一般方法应用于此，基于一种可解释的参数化来确立光滑性，同时也得出与IV型模型相关的自由度数量。

定理10。 假设离散分布严格为正，则链图的*IV*型模型是分层的边际对数线性模型，因此是光滑的。如果链图的分量 $\mathcal{K}_1, \ldots, \mathcal{K}_T$ 编号良好，则该参数化基于以下边缘分布

$$\{\mathrm{PA}(\mathcal{K}_t) \cup \mathcal{X} : \mathcal{X} \subseteq \mathcal{K}_t\}^*, \mathcal{K}_1 \cup \ldots \cup \mathcal{K}_t,\ t = 1, \ldots, T, \tag{33}$$

其中，$\{\ \}^*$ 表示集合中元素的一个非递减次序。将以下效应对应的参数设为零，即可定义该模型

$$\{\mathbb{ID}(\mathcal{X}, \mathcal{K}_t \setminus \mathcal{X} \setminus \mathrm{nb}(\mathcal{X}), \mathrm{PA}(\mathcal{K}_t)) : \mathcal{X} \subseteq \mathcal{K}_t\} \cup \tag{34}$$

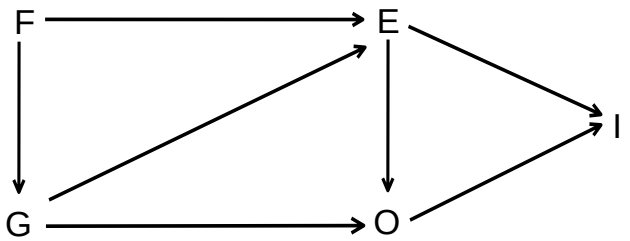
$$\{\mathbb{ID}(\mathcal{X}, \mathrm{PA}(\mathcal{K}_t) \setminus \mathrm{pa}(\mathcal{X}), \mathrm{pa}(\mathcal{X})) : \mathcal{X} \subseteq \mathcal{K}_t\} \cup \mathbb{ID}(\mathcal{K}_t, \mathrm{PRE}(\mathcal{K}_t) \setminus \mathrm{PA}(\mathcal{K}_t), \mathrm{PA}(\mathcal{K}_t)),$$

对于 所有分量 $\mathcal{K}_t$，其中$\mathrm{PRE}(\mathcal{K}_t)$ 是先于 $\mathcal{K}$ 的分量集合 $_t$.

证明见 Rudas 等（2010）。未设为零的参数，即不对应于（34）的参数，构成了该模型的参数化。这些参数与 Marchetti 和 Lupparelli（2008年）在其研究的示例中发现取非零值的效应相对应，尽管用于参数化的边缘分布不同。

其他相关研究包括 Marchetti 和 Lupparelli（2011年），他们描述了多元回归类型的链图模型的边际对数线性参数化。Evans 和 Richardson（2013年）提出了一类与无环有向混合图（ADMG）对应的边际模型，这类图同时包含有向边和双向边。这些

图3：路径模型的图示



研究表明，这些模型具有光滑参数化，并给出了使该参数化具有变异独立性的条件。Nicolussi 和Colombi（2017年）研究了II型链图模型。一般而言，已知这类模型并不光滑，但通过使用边际对数线性参数化，可以确定一个光滑子类。

9.2 路径模型

路径模型在统计学中有着悠久的历史，其基本思想可用图3加以说明。直观地说，当变量 F 影响 E 和 G、 G 影响 E 和 O、 E 影响 O 和 I，以及 O 影响 I时，人们可能希望使用路径模型来描述这种情况。然而，在这种情况下，人们可以说，除了 F 对 E的直接影响之外， F 还通过 G对 E 产生间接影响。同样， E 直接且间接地影响 I 。此外，人们可能希望假设，变量之间除了图中所描绘的影响之外，不存在其他影响。这里还有一个进一步的假设，这一假设经常被采用，但通常仍未明说，即未被纳入考虑的变量对所分析变量的影响都可以忽略不计。路径分析旨在精确地表述这些假设，同时量化这些影响的大小。

为实现这些目标，受Goodman（1973年）的启发，Rudas 等人（2006）提出了以下两步方法。

首先，将图3中的图解释为图形马尔可夫模型，并将此模型中的分布参数化为边际对数线性模型。

与该图相关的条件独立性为

$$O \perp\!\!\!\perp F|GE$$

and

$$I \perp\!\!\!\perp FG|EO.$$

这些条件独立性可以方便地在基于边缘分布的边际对数线性模型中加以施加

$$FGEO \text{ and } FGEOI,$$

并且，正如定理9所示，可通过将以下边际对数线性参数设为零来获得

$$\lambda^{FGEO}_{FO}, \lambda^{FGEO}_{FEO}, \lambda^{FGEO}_{FGO}, \lambda^{FGEO}_{FGEO},$$

[7]变量的记号将在稍后澄清。

在 $FGEO$ 边际分布中，同时

$$\lambda_{FI}^{FGEOI}, \lambda_{GI}^{FGEOI}, \lambda_{FGI}^{FGEOI}, \lambda_{FEI}^{FGEOI}, \lambda_{FOI}^{FGEOI}, \lambda_{GEI}^{FGEOI},$$

$$\lambda_{GOI}^{FGEOI}, \lambda_{FGEI}^{FGEOI}, \lambda_{FGOI}^{FGEOI}, \lambda_{FEOI}^{FGEOI}, \lambda_{GEOI}^{FGEOI}, \lambda_{FGEOI}^{FGEOI}$$

在 $FGEOI$ 边际分布中，也就是在全表中。

总共有 $2^5 = 32$ 个参数，其中16个被设为零，以表示条件独立性。其余参数用于参数化马尔可夫模型中的分布。在本例中，这些参数被解释为衡量影响的强度，而非关联，因为用有向图来表述研究假设，本身就蕴含了这一假设。

其余参数根据其效应中所包含变量的数量，分为四个互不相交的组：空效应的边际对数线性参数；效应中包含一个变量的边际对数线性参数；效应中包含两个变量的边际对数线性参数；以及效应中包含两个以上变量的边际对数线性参数。

效应中包含两个以上变量的参数，用于量化多个变量对某一变量的联合影响，因为这些参数是对数线性参数（在特定边际分布中确定），并具有对数线性参数的标准性质。例如， λ_{FGE}^{FGEO}未被设为零，它衡量了 F 和 G 对E的联合影响，以及它们各自的影响。

尽管路径模型的意图是假设此类高阶影响不存在，但其存在性尚未被排除。事实上，很容易找到这样的分布：例如，三个分类变量之间不存在个体影响（所有二元边际分布均为均匀分布），但其中两个变量共同决定第三个变量，这说明除个体影响之外，确实还存在联合影响（例如参见 Rudas，2018年）。

因此，在路径模型定义的下一步中，将排除这类高阶交互作用。

其次，假设第一步中未设为零的边际对数线性参数里，效应涉及两个以上变量的参数均等于零。

在这个例子中，这意味着将以下边际对数线性参数设为零

$$\lambda_{FGE}^{FGEO}, \lambda_{GEO}^{FGEO},$$

and

$$\lambda_{EOI}^{FGEOI}.$$

要从图模型定义路径模型，还需将另外三个参数设为零。这意味着排除 F 和 G 对 E 的联合影响，以及 G 和 E 对 O 在 $FGEO$ 边际分布中的联合影响，也排除 E 和 O 对 I 在 $FGEOI$ 边际分布中的联合影响。

其余的 $32 - (16 + 3) = 13$ 边际对数线性参数对图3中图所对应的路径模型中的所有分布进行参数化。这些参数包括变量的单变量分布，以及与图3中箭头相关的影响强度。

边缘分布	$FGEO$	$FGEOI$
效应	$\emptyset$, F, G,E, O, FG, FE,FO, GE, GO, EO, FGO, FGE, FEO, GEO, $FGEO$	I, FI,GI,EI, OI,FGI, FEI, FOI , GEI, GOI, EOI, $FGEI$, $FGOI$, $FEOI$, $GEOI$, $FGEOI$
效应设为零以定义图模型	FO,FEO,FGO, $FGEO$	FI, GI, FGI, FEI, FOI, GEI, GOI, $FGEI$, $FGOI$, $FEOI$、$GEOI$、$FGEOI$
效应设为零以定义路径模型	FGE, GEO	EOI
剩余的效应用于参数化路径模型	$\emptyset$, F, G, E, O, FG, FE,GE,GO, EO	I, EI, OI

表5：de　　与图3相关的路径模型的定义和参数化

模型定义与参数化的步骤总结于表5。需要指出的是，这些参数为路径模型中的所有分布提供了参数化。在实际的数据分析情境中，这意味着，如果采用某个特定的路径模型，那么数据中的所有相关信息都由从数据中得到的这些参数估计值加以概括。N´emeth 和 Rudas（2013a）以社会地位获得为背景，提供了一个这样的例子，其中涉及的变量包括: F 父亲的教育程度、 G 父亲的职业、 E 儿子的教育程度、 O 儿子的职业，以及 I 儿子的收入。他们发现，与图3中的图相对应的路径模型能够很好地拟合多个国家的数据，并给出了该模型参数的估计值及其解释。有关边际模型在社会流动研究中的应用详情，请参阅 N´emeth 和 Rudas（2013b）。

9.3 潜变量模型

当分析中的某些相关变量无法观测（即为潜变量）时，对观测变量的分析适用于整个列联表的边际分布。因此，潜变量模型与边际模型密切相关。在某些建模假设下，可以估计潜变量和观测变量的联合分布，但即使如此，模型检验也必须局限于比较估计的边际分布与观测到的边际分布。

例如，如果某人在左右政治尺度上的真实位置，比如 X，难以观测，那么可以提出两个相关的问题，比如 A 和 B。这两个问题是 X 的指标变量，但可能无法对其进行精确测量，而是带有一定的测量误差。因此，如果 A 和 B 等于 X，并受到与 X 无关、且彼此独立的测量误差干扰，那么就有了

$$A \perp\!\!\!\perp B|X. \tag{35}$$

当观测值为数值时，通常假定测量误差是加性的。对于分类数据，测量误差也可能有不同的形式。例如，对于二元变量，测量误差可能以一定的概率改变其类别。如果改变的概率

不取决于真实类别，则误差是独立的。对于有多个类别的变量，独立误差也可能改变其类别，使得报告任何非真实类别的概率都相同，且不取决于真实类别。

由于 X 是潜变量，而 A 和 B 是观测变量，(35)是一个简单的潜变量模型。如上所述，它有许多直接明了的边际对数线性模型定义。可以使用 X 和 XAB 边缘分布，或 X、 A、 B、 XAB 边缘分布，但定义也可以基于 XAB 全表。在这两种情况下，模型定义为

$$\lambda_{AB}^{XAB} = \lambda_{XAB}^{XAB} = 0.$$

要检验潜变量模型(35)，必须依赖 AB边际分布的观测数据。通常的做法是指定潜变量X的类别数，并在满足(35)的条件下估计分布 AB ，使观测数据的似然最大化。通常采用EM算法来求得此类估计值（例如参见Rudas，2018年）。然后使用某种统计检验，将估计值与实际观测值进行比较。

Bergsma等人（2009）、Bergsma、Croon和Hagenaars（2013年）以及Hagenaars、Bergsma和Croon（2019）介绍了边际建模方法在潜变量模型中的若干更复杂应用。在其中一个问题中，有两个相关的潜变量， Y和 Z。他们的例子涉及选举预测，其中 Y 是三个政党之一的政党偏好， Z 则是对各自候选人的候选人偏好。这些被视为潜变量，可能只能以不精确的方式观测到。观测变量是面板研究两个调查波次中，针对政党偏好（A 和 B）以及候选人偏好（C 和 D）问题的回答。这一设置可视为第2.1节所讨论的重复测量设计的一个实例。他们拟合的模型是一种路径分析类型的图模型，其含义是，允许的最高阶交互项为 YZ, YA, YB, ZC, ZD。

要给出该模型的边际对数线性定义，可以使用边缘分布YZ， XA， XB， ZC， ZD, $YZABCD$，并将 $YZABCD$ 边际分布中定义的所有边际对数线性参数设为零。

对 Y 和 Z 的单变量边际分布的估计显示了各政党及其 respective 候选人的相对受欢迎程度。Bergsma等人（2009）进一步研究了这些边际分布相同的假设，即候选人与提名他们的政党同样受欢迎。该假设称为潜在边际齐性。

为了表述潜在边际齐性，最简单的做法是使用以下边缘分布：Y， Z， YZ， YA, YB， ZC， ZD， $YZABCD$。这里同样，通过将所有在 $YZABCD$ 边际分布中计算得出的参数设为零来施加路径模型，而潜在边际齐性则通过要求.....来施加。

$$\lambda_{Y}^{Y} = \lambda_{Z}^{Z},$$

这是一个边际对数线性模型。

显变量通常被视为潜变量的指标变量。此类指标变量的信度是指显变量由潜变量决定的程度。当然，这可以通过多种方式进行测量；其中一种方式是基于给定潜变量条件下显变量的条件分布。

为考虑一个非常简单的模型，设 Y 和 Z 是潜变量，分别具有显变量指标 A 和 C 。我们现在并不关心潜变量或显变量之间是如何关联的，而只关心 Y 在多大程度上决定 A ，以及 Z 在多大程度上决定C。根据定理 2，如果在边际对数线性参数化中使用 YZ、YA 和 ZC 边缘分布，那么λ_A^{YA}和 λ_{YA}^{YA} 决定给定 Y时 A 的条件分布，而λ_C^{ZC}和 λ_{ZC}^{ZC} 决定给定 Z 时 C 的条件分布。

如果现在假设所有变量都具有相同的类别，例如上例中的政党和候选人偏好，那么这一要求是：

$$\lambda_A^{YA} = \lambda_C^{ZC} \text{ and } \lambda_{YA}^{YA} = \lambda_{ZC}^{ZC},$$

表示两个显变量的信度相等。

这种分析信度的方法的优势在于，只要相关边际分布可以按非递减顺序排列，就能将它与任何其他建模假设结合起来。详情及应用请参见 Bergsma 等人（2009）。

在捕获-再捕获模型的背景下，也有人研究了带有潜变量的边际对数线性模型（Stanghellini 和 van der Heijden（2004年）、Bartoluci和 Forcina（2006））。在这种情况下，观测变量在给定潜变量的条件下也不一定相互独立。

9.4 进一步的应用与扩展

本节简要概述了文献中发表的一些进一步的理论发展和有趣的应用。

Qaqish 和 Ivanova（2006）研究了多变量逻辑参数化，这些参数化可由上述定义的边际对数线性参数化加以推广，并给出了此类参数强相容性的相关结果。

Bartoluci 和 Forcina（2006）将边际对数线性参数化应用于捕获-再捕获问题的模型构建。相关研究还可参见 Turner（2007）。

Forcina（2008）构建了含协变量的潜在类别模型的边际对数线性参数化，并得出了可识别性方面的结果。

Bartolucci 等人（2012）提出了一种贝叶斯方法，用于从一大类边际模型中选出数据支持度最高的模型；这类模型由广义对数几率或广义优势比上的等式或不等式约束定义。他们使用贝叶斯因子来决定模型选择。

Dardanoni 等人（2012）分析了多个国家的代际社会经济流动表，以检验单调性假设：较高的社会经济阶层绝不会比其下的阶层处于劣势。他们使用 Bartolucci 等人（2007）提出的参数化，将这一单调性表述为边际模型。

Shpitser 等人（2013年）为嵌套马尔可夫模型发展了边际对数线性参数化，并施加了类似于第9.2节所述思想的稀疏性。

Kuijpers等人（2013a）提出了将广泛使用的测试分数信度指标 Cronbach's alpha 表述为边际对数线性模型，并据此检验假设的方法。Kuijpers等人（2013b）采用边际建模方法，给出了题目非二元时以及题目数量较多时可扩展性系数的标准误。

Colombi 和 Giordano（2015）用边际对数线性参数对潜在马尔可夫模型的两个分量（观测到的关系序列和未观测到的马尔可夫链）进行参数化，并表明将其中一些参数设为零即可表述相关假设。

Colombi 和 Forcina（2016）检验边际对数线性参数的不等式假设。他们提出一种基于似然的程序，用于检验一组等式约束是否存在正向偏离（即不等式约束），随后再检验后者与饱和模型之间的差异。

Ntzoufras 等人（2019）探讨图形边际对数线性模型中贝叶斯推断的若干方面。他们提出一种执行马尔可夫链蒙特卡洛的方法，以获得后验密度。他们的方法还考虑了参数值应以能够确保边际分布相容的方式选取这一要求。

Colombi 等人（2019）用一个二元变量对评分者的潜在行为建模，该变量表示两种可能策略之一。他们采用边际参数化，将响应与潜在解释因素联系起来。

Bon 等人（2020）研究敏感或机密数据的披露限制。他们将数据保管方提供的部分信息建模为超表可能重叠的边缘分布上的对数线性模型，并探讨整合现有信息的方法。他们还将该方法应用于澳大利亚住房产权转换数据。

Nicolussi 和 Cazzaro（2020年）分析了情境特定的独立性，即仅在所涉及变量的某些类别组合中成立、而非在所有类别组合中都成立的独立性，并指出层级多项式边际模型可用于描述此类关系。

Roverato、Lupparelli 和 La Rocca（2013年），以及 Lupparelli 和 Roverato（2017年）研究了二元数据边际模型的对数均值线性 参数化。这些参数化是本文所研究的边际对数线性参数化的替代方案。对数均值线性参数化有一个有趣的优点：其似然具有闭式解。

Bergsma、Croon 和 Hagenaars（2013年）展示了如何将边际建模方法扩展至复杂抽样设计；具体而言，他们分析了一组通过旋转面板设计收集的数据。该研究的分析基于具有部分依赖性的数据。此外，他们还展示了如何将边际建模用于复杂统计模型，并举例说明如何使用潜变量，以及对单元概率施加对数线性和非对数线性约束来进行数据分析。

参考文献

[1] Agresti, A.: 分类数据分析，第*3*版。 Wiley（2013年）[2] Andersson, S. A., Madigan, D., Perlman, M. D. 链图的替代马尔可夫性质。斯堪的纳维亚统计学杂志，*28*, 33-85（2001年）。[3] Aitchison, J. 和 Silvey, S. D.：受约束参数的最大似然估计。数学统计年刊，*29*，813-828（1958年）。[4] Barndorff-Nielsen, O.（1978）信息与指数族。纽约：Wiley。

[5] Bartolucci, F., Forcina, A.：含连续协变量的捕获–再捕获数据潜在边际模型类。美国统计学会期刊 101786–794（2006）[6] Bartolucci, F., Colombi, R., Forcina, A.：一种扩展的边际链接函数类，用于通过等式与不等式约束对列联表建模。中华统计学志691–711（2007年）[7] Bartolucci, F., Scaccia, L., Farcomeni, A.：通过包容先验和重要性抽样，对一类分类数据边际模型进行贝叶斯推断。计算统计与数据分析 56 4067-4080（2012）Bergsma, W., Rudas, T.：分类数据的边际模型。统计年刊，30，140-159（2002a）[8] Bergsma, W., Rudas, T.：多变量分类分布的变异独立参数化。载于：Cuadras, C. M., Fortiana, J., Rodriguez-Lallena, J. A.（编）给定边际分布与统计建模 Kluwer，21-27。（2002b）[9] Bergsma, W., Rudas, T.：论条件关联与边际关联。图卢兹科学学院年刊，6 11，455-468（2003年）[10] Bergsma, W., Croon, M., Hagenaars, J.A.：相关、聚类和纵向分类数据的边际模型。施普林格，纽约（2009）[11] Bergsma, W. P., Croon, M. A., Hagenaars, J. A.（2013年）。分类数据边际建模的进展。社会学方法论，43（1），141。[12] Bergsma, W. P. 和 Rapcs´ak, T.（2005）一种用于光滑等式约束优化的精确罚函数法及其在最大似然估计中的应用。*Eurandom* 技术报告。[13] Bishop, Y. M. M., Fienberg, S. E., Holland, P. W.离散多变量分析：理论与实践。 麻省理工学院出版社（1975年）[14]Bon, J., Baffour, B., Spallek, M., Haynes, M.：分析动态生成的重叠列联表中的敏感数据。官方统计学期刊36，275-296（2020年）[15] Cocchi, M.（编）：数据融合方法论与应用。 爱思唯尔（2019）[16] Colombi, R., Forcina, A.：一类满足边际独立性和上下文特定条件独立性的光滑模型。多变量分析期刊，126，75-85（2014）[17] Colombi, R., Forcina, A.：检验列联表中的序约束。*Metrika*，79，73-90（2016）[18] Colombi, R. 和 Forcina, A.（2001年）。用于分析有序响应变量正关联的边际回归模型。*Biometrika*，88，1007-1019。

[19] Colombi, R., Giordano, S.：用于分类时间序列的多个隐马尔可夫模型。《多变量分析期刊》，140，19-30（2015）[20] Colombi, R., Giordano, S., and Cazzaro, M.（2014）。 hmmm：用于层级多项式边际模型的 R 软件包。《统计软件期刊》，59(11)，1–25。[21] Colombi, R., Giordano, S., Gottard, A., Iannario：具有潜在不确定性的层级边际模型。《斯堪的纳维亚统计学杂志》，46，595-620（2019）[22] Cox, D. R. and Wermuth, N.（1996）。《多变量依赖关系》。伦敦：Chapman and Hall。[23] Dardanoni, V., Fiorini, M., Forcina：代际流动表中的随机单调性。《应用计量经济学期刊》，27，85-107（2012）[24] D'Orazio, M., Di Zio, M., Scanu：《分类数据的统计匹配：呈现不确定性并使用逻辑约束》。《官方统计学期刊》，22，137-157（2006）[25] Drton, M.（2009）。离散链图模型。《Bernoulli》，15(3)，736-753。[26] Eppmann, H., Kr¨ugener, S., Sch¨afer, J.：德国首次基于登记资料的人口普查（2011年）。《Allgemeines Statistisches Archiv》，90(3)，465-482（2006）[27] Evans, R. J.：边际对数线性参数化的光滑性。《电子统计学期刊》，9，475-491（2015）[28] Evans, R. J. and Forcina, A.：拟合受约束边际模型的两种算法。《计算统计与数据分析》，66：1–7，2013年。[29] Evans, R. J. and Richardson, T. S.（2013）。图形马尔可夫模型的边际对数线性参数。《英国皇家学会会刊 B 辑：统计方法论》，75(4)，743。[30] Forcina, A.：含个体协变量的扩展潜在类别模型的可识别性。《计算统计与数据分析》，52，5263-5268（2008）[31] Forcina, A.：条件独立性的光滑性。《多变量分析期刊》，106，49-56（2012）[32] Forcina, A. and Kateri, M.：一类新的广义 RC 关联模型：估计与主要性质。《多变量分析期刊》，184，1-16（2021）[33] Forcina, A., Lupparelli, M., Marchetti, G. M.：由一组条件独立性定义的离散模型的边际参数化。《多变量分析期刊》101 2519-2527（2010）[34] Frees, E. W, Kim, J.-S.：面板研究。载于 Rudas, T.（编），《概率论手册：理论与应用》，205-224，Sage，千橡市（2008）

[35] Frydenberg, M.（1990）。链图马尔可夫性质。《斯堪的纳维亚统计学杂志》，17，333-353。[36] Ghosh, S., Vellaisamy, P.：分类数据的边际对数线性参数及其可塌缩性。arXiv 1711.00680v4（2019）[37] Glonek, G. F. V.：多元分类响应的回归模型类。《Biometrika》，83，15-28（1996）[38] Glonek, G. F. V., McCullagh, P.：多元逻辑回归模型。《英国皇家学会会刊 B 辑》，57，533-546（1995）[39] Goodman, L. A.：当某些变量具有时间先后关系时多维列联表的分析：一种修正的路径分析方法。《Biometrika》，60，179-192（1973）[40] Hagenaars, J. A., Bergsma, W., and Croon, M.（2019）。非对数线性边际潜在类别模型。载于：《潜在类别分析进展：献给 C. Mitchell Dayton 的纪念文集》，61。[41] Kuijpers, R. E., Ark, L. A., Croon, M. A.：使用边际模型检验涉及 Cronbach α 的假设。《英国数学与统计心理学期刊》，66，503-520（2013a）[42] Kuijpers, R. E., Ark, L. A., Croon, M. A.：使用边际模型进行 Mokken 量表分析时可扩展性系数的标准误和置信区间。《社会学方法论》，43，42-69（2013b）[43] Lang, J. B.（1996a）。广义对数线性模型类的最大似然方法。《统计年刊》，24，726-752。[44] Lang, J. B.（1996b）。多项式与泊松对数线性模型的比较。《英国皇家学会会刊：B辑（方法论）》，58(1)，253-266。[45] Lang, J. B. and Agresti, A.（1994）。同时对多元分类响应的联合分布和边际分布进行建模。《美国统计学会期刊》，89，625-632。[46] Lauritzen, S.L.：《图模型》。Clarendon Press，牛津（1996）[47] Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H.-G.（1990）。有向马尔可夫场的独立性性质。《Networks》，20(5)，491–505。[48] Lauritzen, S. L. and Wermuth, N.（1989）。定性变量与定量变量并存时变量间关联的图模型。《统计年刊》，17，31-57。[49] Liang, K. Y. and Zeger, S. L.（1996）。使用广义线性模型分析纵向数据。《Biometrika》，73，13-22。

[50] Lipsitz, S. R., Laird, N. M., 和 Harrington, D. P.（1991）。相关二元数据的广义估计方程：使用优势比作为关联的度量。Biometrika，78(1)，153-160。[51] Little, R., Rubin, D.：《缺失数据统计分析》，第3版。Wiley，纽约（2019）[52] Lupparelli, M., Marchetti, G. M. 和 Bergsma, W. P.（2009）。将双向图模型参数化并拟合至分类数据。《斯堪的纳维亚统计学杂志》，36(3)，559–576。[53] Lupparelli, M., Roverato, A. 二元响应的对数均值线性回归模型及其在多病共存研究中的应用（2017）。《英国皇家统计学会杂志：C辑（应用统计）》，66(2)，227–252。[54] Marchetti, G. M. 和 Lupparelli, M.（2011）分类数据的多元回归型链图模型。《Bernoulli》，17(3)，827–844。[55] McCullagh, P., Nelder, J. A.：《广义线性模型》，第2版，Chapman and Hall，伦敦（1989年）[56] Molenberghs, G. 和 Verbeke, G.：《离散纵向数据模型》（2005）[57] N´emeth, R., Rudas, T.：关于离散边际图模型的应用。《社会学方法论》，43，70-100（2013a）[58] N´emeth, R., Rudas, T.：社会流动研究中的离散图模型——对国家社会主义崩溃前美国、捷克斯洛伐克和匈牙利社会流动的比较分析。《社会学方法论通报》，118，5-21（2013b）[59] Nicolussi, F., Cazzaro, M.：层级多项式边际模型中的情境特定独立性。《统计方法与应用》，29，767-786（2020）[60] Nicolussi, F., Colombi, R.：分类数据的II型链图模型：一个光滑子类。《Bernoulli》，23，863-883（2017）[61] Ntzoufras, I., Tarantola, C., Lupparelli, M.：用于图形对数线性边际模型贝叶斯定量学习的基于概率的独立抽样器。《贝叶斯分析》，14，777-803（2019）[62] Pan, W.（2001）。广义估计方程中的赤池信息准则。《Biometrics》，57(1)，120-125。[63] Qaqish, B. F., Ivanova, T.：《多元逻辑回归模型》。《Biometrika》，93，1011-1017（2006）[64]Rhemtulla, M., Little, T.：研究认知发展的工具：计划性缺失数据设计。《认知与发展杂志》，13.4，10.1080/15248372.2012.717340（2012）

[65] Rotnitzky, A., 和 Jewell, N. P.（1990）。针对簇相关数据的半参数广义线性模型中回归参数的假设检验。*Biometrika*，77(3)，485-497。[66] Roverato, A., Lupparelli, M., La Rocca, L.（2013）。二元数据的对数均值线性模型。*Biometrika*，(2)，485–494。[67] Rudas, T.：预设条件交互结构模型及其在流动性表分析中的应用。*Quality and Quantity* 25，345-358（1991）[68] Rudas, T.：列联表分析中的优势比。第119号，《社会科学定量应用》，Sage（1998）[69]Rudas, T.：信息性分配与一致性处理选择。统计方法论，7，323-337（2010）[70] Rudas, T.：多元二元分布的方向可折叠参数化。统计方法论，27，132-145（2015）[71] Rudas, T.：分类数据分析讲义。施普林格，纽约（2018）[72]Rudas, T., Bergsma, W.：边际模型在分类数据中的应用。*Metron*，**42**，15-37（2004）[73] Rudas, T., Bergsma, W., N´emeth, R.：分类数据路径模型的参数化与估计。载于 Rizzi, A., Vichi, M.（编）*COMPSTAT 2006*，383-394，Physica Verlag，海德堡（2006）[74] Rudas, T., Bergsma, W., N´emeth, R.：条件独立模型的边际对数线性参数化。*Biometrika*，**97**，1006-1012（2010）[75] Rudas, T., Leimer, H.-G.：已知条件优势比或已知对数线性参数的列联表分析。载于：Francis, B., Seeberg, G. U. H., van der Heijden, P. G. M., Jansen, W.（编）统计建模 313-322，爱思唯尔（1992）[76] Shpitser, I., Evans, R. J., Richardson, T. S., Robins, J. M.：具有对数线性参数的稀疏嵌套马尔可夫模型。载于：第二十九届人工智能不确定性会议，576-585。2013年[77] Stanghellini, E., van der Heijden, P. G. 一种考虑已观测和未观测异质性的多记录系统估计方法。*Biometrics*，*60*(2)，510–516（2004）[78] Touloumis, A., Agresti, A., 和 Kateri, M.（2013）。使用局部优势比参数化的多项响应广义估计方程。*Biometrics*，69(3)，633-640。[79] Turner, E. L.：捕获-再捕获数据的边际建模。博士论文。麦吉尔大学 蒙特利尔（2007）