二元贝塔分布：参数推断与诊断方法

Lucas Machado Moschen、Luiz Max Carvalho，应用数学学院，Fundação Getulio Vargas。

2023年3月2日

arXiv:2303.01271v1 [stat.ME] 2 Mar 2023

摘要

相关比例广泛存在于现实应用中。由于其受约束的特性，要为其找到合适的概率模型是一项独特的挑战。双变量贝塔分布是对广为人知的贝塔分布的一种自然扩展，可用于描述 $[0,1]^2$ 上的相关量。不过，其构造并非唯一。多年来，人们提出了许多双变量贝塔分布，其参数数量从三个到八个甚至更多不等，且联合密度和分布矩在数学可处理性方面各有差异。本文研究 Olkin 和 Trikalinos (2015) 提出的构造，该构造在参数丰富度与可处理性之间取得了平衡。我们分别采用矩估计法以及结合哈密顿蒙特卡洛的数据增广与潜变量方法，介绍经典（频率学派）和贝叶斯估计方法。本文还讨论了将双变量贝塔分布用作先验分布时的 elicitation。借助蒙特卡洛实验，我们深入探讨了用于检验模型拟合与适切性的诊断方法，实验涵盖了数据生成设定正确和设定错误两种情形。

关键词：双变量贝塔分布；相关比例；诊断方法；矩估计法；贝叶斯 an 估计。

1 引言

相关比例出现在许多现实世界的应用中，例如森林中的花粉分布建模（Nadarajah 等人，2017）、干旱频率与持续时间之间的关系（Nadarajah 等人，2007），以及不完美检测系统的灵敏度/特异度（例如疾病检测，Dahabreh 等人（2013））。贝塔分布是一维随机变量广泛使用的分布，其支撑集为 $[0,1],$，而向平方区域 $[0,1]\times[0,1]$ 的扩展是自然的，由此得到双变量贝塔分布，即具有服从贝塔分布的边缘分布的二元分布。可采用许多构造方法，其差异取决于所需参数的数量、可达到的相关结构，以及联合概率密度函数（pdf，见下文）在数学和计算上的可处理性。

Balakrishnan 和 Lai（2009）对连续双变量分布进行了全面综述，特别关注具有贝塔边缘分布的分布，涵盖从狄利克雷分布到通过连接函数进行的构造。Trick 等人（2021）对该主题进行了更新评估，重点关注通过服从伽马分布的随机变量变换构造的分布。涉及双变量贝塔分布的论文包括 Libby 和 Novick（1982）、Olkin and Liu（2003）、Magnussen（2004）、Nadarajah and Kotz（2005）、Sarabia 和 Castillo（2006）、Nadarajah 等人（2007）、Arnold 和 Ng（2011）、Nadarajah 等人（2017）以及 Trick 等人（2021），但这并非完整列表。

引入贝塔边缘分布之间相关性的一种简单且常用的方法，是使用共享的伽马分布随机变量。Libby 和 Novick (1982) 讨论了一种多变量分布，其边缘分布为具有三个参数的广义贝塔分布，联合密度具有闭式表达式，而矩则通过级数表示。在双变量情形下，这是一个六参数分布。 Olkin and Liu(2003) 提出了这种分布的一个特例，仅使用三个参数。 Sarabia 和 Castillo (2006) 通过研究若干边缘分布服从广义贝塔分布的双变量分布，拓展了这一结果。他们认为，需使用三参数分布才能对边缘分布的均值、方差和偏度进行建模。 Nadarajah 等人 (2007) 构造了另一种二元贝塔分布，其边缘分布为广义贝塔分布，但其矩没有闭式表达式，且变量模拟需要采用拒绝抽样。另一方面，其联合密度除去一个依赖于参数的常数因子后，具有解析表达式。 Nadarajah 等人 (2007) 随后为他们的双变量贝塔构造提出了最大似然估计量和矩估计量。

Arnold 和 Ng （2011）提出了一种由具有共同尺度参数的独立伽马分布构成的五参数模型。该分布的一个优点是边缘之间的相关性可以覆盖整个范围（−1，1），这使其有别于先前的构造。

然而，密度和乘积矩无法用闭式形式表示。因此，Arnold 和 Ng （2011）采用了一种修正的最大似然估计量：通过最大化边缘似然来估计边缘参数，并使用矩估计法估计相关参数。该分布存在两个问题：（i）矩估计法得出的参数估计值可能为负；一种启发式解是在这种情况下将其设为 0；（ii）它不支持任意贝塔边际分布。作者通过蒙特卡罗模拟研究评估了估计过程中的所有变体。Arnold 和 Ng Arnold 和 Ng （2011）提出的另一个问题——Olkin 和 Trikalinos（2015）对此进行了讨论——是难以将该分布扩展到更高维度。

Nadarajah 等人（2017）提出了一种具有初等概率密度函数的六参数分布，并通过数值方法论证相关系数覆盖范围 $[0,1)$，尽管他们未给出数学证明。乘积矩具有双重无穷级数表示，其边缘分布是贝塔分布的推广。参数估计采用最大似然法。特别是，作者将其与其他具有初等概率密度函数的二元贝塔分布进行了比较，例如 Libby 和 Novick （1982）、Sarabia 和 Castillo （2006）以及Nadarajah 等人（2007）。最近，Trick 等人（2021）研究了一种六参数二元贝塔分布，扩展了Magnussen （2004）的工作。该分布对任意贝塔边际分布进行建模，但其缺点是仅允许正相关，且乘积矩无法以闭式形式获得。作者指出，不同的两组参数设定会生成相似的数据，这在实际应用中会导致可识别性问题。为了解决参数推断问题，他们将该分布简化为五参数分布，并对参数空间施加约束。

Barros （2015）提出了一种针对 Nadarajah andKotz（2005）所提出的双变量贝塔分布的估计方法。该方法基于这样一个事实：相互独立的贝塔分布随机变量的乘积仍服从贝塔分布。在这种建模方式下，矩可以写成闭式形式，但联合概率密度函数则不能。此外，此情形下的相关性严格为正。 Crackel 和 Flegal（2017）提出了一种贝叶斯方法，用于估计Arnold 和 Ng （2011）分布的参数，并将原始模型扩展为包含八个参数。由于似然难以处理，他们使用近似贝叶斯计算（ABC）算法来近似后验分布。

前文的介绍清楚地表明，我们需要在参数丰富度和可处理性之间取得平衡：例如，有些构造包含许多参数且具有灵活性，因而能够涵盖完整范围的相关系数。然而，在许多情况下，联合密度难以处理，从而使基于似然的方法变得复杂。另一些情况下，矩无法写成闭式形式，因此难以采用高效的矩估计技术，也不易直观解释参数对矩的影响。下文将详细介绍 Olkin 和 Trikalinos （2015）提出的构造；我们认为，该构造在可处理性与参数丰富度之间取得了良好的平衡。

Olkin 和 Trikalinos（2015）描述了一种边缘分布服从贝塔分布的二元分布，其在空间 $(0,1)\times(0,1)$ 上具有正概率，并且相关性覆盖整个范围 $(-1,1)$。其构造如下：令 $\boldsymbol{U}\sim$ 狄利克雷$(\boldsymbol{\alpha})$，其中 $\boldsymbol{U}=(U_1,U_2,U_3,U_4)$ 且 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$，使得 $\alpha_i>0$for $i=1,\ldots,4$ 且 $U_4=1-U_1-U_2-U_3$。 $\boldsymbol{U}$相对于 $\mathbb{R}^3$ 上的勒贝格测度的联合密度为

$$f_U(u_1,u_2,u_3)=\frac{1}{B(\boldsymbol{\alpha})}u_1^{\alpha_1-1}u_2^{\alpha_2-1}u_3^{\alpha_3-1}(1-u_1-u_2-u_3)^{\alpha_4-1}, \tag{1}$$

当 $u_i\in[0,1]$ for $i=1,2,3$ 且 $u_1+u_2+u_3\leq 1$ 时。否则， $f_U(u_1,u_2,u_3)=0$。归一化常数定义为

$$B(\boldsymbol{v})=\frac{\prod_{i=1}^n\Gamma(v_i)}{\Gamma\left(\sum_{i=1}^n v_i\right)},$$

其中 $\boldsymbol{v}\in\mathbb{R}^n$ 具有正元素。

令 $X=U_1+U_2$和 $Y=U_1+U_3$。在第 2 节中，我们表明 X和 Y的边缘分布为贝塔分布，因此二元随机向量（X，Y）具有参数为$\boldsymbol{\alpha}$的二元贝塔分布。这是一种四参数构造，其边缘分布之间的相关系数范围完整，且其联合密度难以处理，但矩具有闭式表达式。此外，随机变量易于模拟，从而便于实施诸如预测等推断任务。

因此，本文将重点放在Olkin 和 Trikalinos （2015）提出的双变量贝塔分布的参数估计上，采用频率学派方法（矩估计法）和贝叶斯方法（结合哈密顿蒙特卡洛的潜变量表示）。我们还讨论推断中一个重要但常被忽视的方面，即开发用于检验模型拟合和适切性的**诊断方法**。

1.1 主要贡献

本文深入研究Olkin 和 Trikalinos （2015）提出的分布的参数估计。具体而言，我们研究在有一个大小为 n的随机样本 $\boldsymbol{Z} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ 可用，且希望估计 $\boldsymbol{\alpha}$时的参数估计。我们的主要贡献有三方面：

- 在矩估计量存在时给出其精确解，确定矩无法产生定义良好的二元贝塔分布的区域，并提出四类基于矩的估计量；
- 使用 Stan 编程语言，通过哈密顿蒙特卡洛进行贝叶斯估计，以近似后验分布 $p(\boldsymbol{\alpha} \mid \boldsymbol{Z})$（Carpenter 等人，2017；Stan 开发团队，2022）。我们提出一种潜变量表示，以规避似然的不可处理性，并实现对后验分布的高效探索。
- 针对频率学派和贝叶斯方法，设计定制诊断方法，以识别模型拟合问题。

2 理论

令 (X, Y) 服从由Olkin 和 Trikalinos (2015)构造的二元贝塔分布，其参数为 $\boldsymbol{\alpha}$。根据聚合性质（Ng 等人，2011，第 2.2 节），

$$(U_1 + U_2, U_3, U_4) \sim \text{Dirichlet}(\alpha_1 + \alpha_2, \alpha_3, \alpha_4),$$

其边缘分布服从贝塔分布，这意味着 $X \sim \text{Beta}(\alpha_1 + \alpha_2, \alpha_3 + \alpha_4)$。同理，$Y \sim \text{Beta}(\alpha_1 + \alpha_3, \alpha_2 + \alpha_4)$，由此可知边缘分布服从贝塔分布。根据这一构造，我们可以计算 X 和 Y 的均值、方差及其他矩。记 $s_\alpha = \sum_{i=1}^4 \alpha_i$，则有

$$\begin{aligned} \mathbb{E}[X] &= \frac{\alpha_1 + \alpha_2}{s_\alpha}, & \mathbb{E}[Y] &= \frac{\alpha_1 + \alpha_3}{s_\alpha}, \\ \text{Var}(X) &= \frac{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)}{s_\alpha^2(s_\alpha + 1)}, & \text{Var}(Y) &= \frac{(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}{s_\alpha^2(s_\alpha + 1)}. \end{aligned} \tag{2}$$

边缘分布的参数之和对于 X 和 Y 相同，这限制了能够生成定义良好的二元贝塔分布的边缘贝塔分布对的族。(X, Y) 的密度见 命题 1；据我们所知，它没有闭式表达式。不过，它可以用特殊函数表示——参见 Olkin 和 Trikalinos (2015) 的附录A。 图 B.16 展示了 $\boldsymbol{\alpha}$ 取不同值时的密度。

命题 1 (二元贝塔密度)。 相对于 $\mathbb{R}^2$ 上的勒贝格测度，(X, Y) 的联合密度为

$$f_{X,Y}(x, y) = \frac{1}{B(\boldsymbol{\alpha})} \int_\Omega u^{\alpha_1 - 1}(x - u)^{\alpha_2 - 1}(y - u)^{\alpha_3 - 1}(1 - x - y + u)^{\alpha_4 - 1}\, du, \tag{3}$$

其中

$$\Omega = (\max(0, x + y - 1), \min(x, y)).$$

参见证明，见附录A 。

正如Olkin 和 Trikalinos （2015）第 2.3 节所述，原点矩可通过以下公式计算

$$\mathbb{E}[X^r Y^s] = \mathbb{E}[(U_1 + U_2)^r (U_1 + U_3)^s].$$

结合（Ng 等人，2011，第 39 页）

$$\text{Cov}(U_i, U_j) = -\frac{\alpha_i \alpha_j}{s_\alpha^2(s_\alpha + 1)}, i, j = 1, \ldots, 4, i \neq j, \tag{4}$$

根据边际矩，我们推导出以下命题：

命题 2（协方差与相关性）。 X 与 Y 之间的协方差为

$$\mathrm{Cov}(X,Y)=\frac{1}{s_\alpha^2(s_\alpha+1)}(\alpha_1\alpha_4-\alpha_2\alpha_3), \tag{5}$$

并且相关性为

$$\mathrm{Cor}(X,Y)=\frac{\alpha_1\alpha_4-\alpha_2\alpha_3}{\sqrt{(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)}}. \tag{6}$$

下面我们证明，表达式（6）使 X 与 Y 之间的相关性可以覆盖整个范围（-1，1）。固定 $\rho\in(-1,1)$，并设 X和 Y的边缘分布分别为 Beta(a，a) $a>0$。要使其具有这些边缘分布的二元贝塔分布，我们需要

$$\alpha_1+\alpha_2=\alpha_3+\alpha_4=\alpha_1+\alpha_3=\alpha_2+\alpha_4=a,$$

其一个解为 $\alpha_1=\alpha_4\in(0,a)$，且 $\alpha_2=\alpha_3=a-\alpha_4$。相关性公式（6）可化简为

$$\mathrm{Cor}(X,Y)=\frac{\alpha_4^2-(a-\alpha_4)^2}{a^2}=\frac{2}{a}\alpha_4-1,$$

这意味着 $\alpha_4=a\dfrac{1+\rho}{2}\in(0,a)$ 是 Cor(X，Y) $=\rho$ 的一个解。因此，参数为 $\boldsymbol{\alpha}=\frac{a}{2}(1+\rho,1-\rho,1-\rho,1+\rho)$ 的双变量贝塔分布的相关性为 ρ。

Ng 等人的定理 2.1 (2011)（第 40 页）阐述了狄利克雷分布与伽马分布之间的关系，由此得到命题 3。这一表示表明，该模型是 Arnold 和 Ng 在 (2011) 第 6.1 节定义的八参数双变量贝塔分布的一个极限情形，此时 $\delta_1,\ldots,\delta_4\to 0$。

命题 3（与伽马分布的关系）。 令 $Y_i\sim$伽马分布(α_i，1) 其中 $i=1,\ldots,4$ 并定义

$$X=\frac{Y_1+Y_2}{Y_1+Y_2+Y_3+Y_4}\quad and\quad Y=\frac{Y_1+Y_3}{Y_1+Y_2+Y_3+Y_4}.$$

则（X，Y）服从参数为 $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_4)$ 的二元贝塔分布。

注1。 假设 $X\in(0,1)$。由于 $1-X>0$，我们有 $\mathbb{E}[X]>\mathbb{E}[X^2]$，这意味着

$$\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2<\mathbb{E}[X]-\mathbb{E}[X]^2=\mathbb{E}[X](1-\mathbb{E}[X]).$$

因此， X 的方差受均值的函数限制。特别地，如果一个变量服从贝塔分布，则它满足上述不等式。

2.1 矩

本节讨论与双变量贝塔分布的矩相关的问题，这些问题对于第3 节和第4节中估计和 elicitation 的处理十分重要。记 $m_1=\mathbb{E}[X]$，$m_2=\mathbb{E}[Y]$，$v_1=$方差(X)，$v_2=$方差(Y) 和 $\rho=$相关系数(X，Y)。如果（X，Y）服从参数为 $\boldsymbol{\alpha}$的二元贝塔分布，则这些量满足以下非线性方程组，我们称之为矩方程组：

$$\begin{cases} m_1=\dfrac{\alpha_1+\alpha_2}{s_\alpha},\\ m_2=\dfrac{\alpha_1+\alpha_3}{s_\alpha},\\ v_1=\dfrac{(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)}{s_\alpha^2(s_\alpha+1)},\\ v_2=\dfrac{(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)}{s_\alpha^2(s_\alpha+1)},\\ \rho=\dfrac{\alpha_1\alpha_4-\alpha_2\alpha_3}{\sqrt{(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)}}. \end{cases} \tag{7}$$

以下命题阐明了（7）中的矩方程组何时存在解 $\boldsymbol{\alpha}$，这意味着存在具有预定矩的二元贝塔分布。具体而言，它指出了何时无法按照 Olkin 和 Trikalinos的方法从矩定义二元贝塔分布，这是模型识别的重要步骤。

命题4 (矩方程组的解)。 方程组 (7) 无解，当

$$\frac{m_1(1-m_1)}{v_1} \neq \frac{m_2(1-m_2)}{v_2}. \tag{8}$$

此外，去掉关于 v_2 的第四个方程后，该方程组有唯一解：

$$\begin{aligned}
\alpha_1 &= (m_1+m_2-1)\bar{\alpha}+\alpha_4,\\
\alpha_2 &= (1-m_2)\bar{\alpha}-\alpha_4,\\
\alpha_3 &= (1-m_1)\bar{\alpha}-\alpha_4,\\
\alpha_4 &= \bar{\alpha}\left(\rho\sqrt{m_1m_2(1-m_1)(1-m_2)}+(1-m_1)(1-m_2)\right),
\end{aligned} \tag{9}$$

其中

$$\bar{\alpha} = \frac{m_1-m_1^2-v_1}{v_1} = s_\alpha. \tag{10}$$

参见附录A中的证明。注意，方程组（7）的前两个方程意味着以下关系

$$\alpha_1 = \frac{m_1+m_2-1}{1-m_1}\alpha_3 + \frac{m_2}{1-m_1}\alpha_4 \tag{11}$$

$$\alpha_2 = \frac{1-m_2}{1-m_1}\alpha_3 + \frac{m_1-m_2}{1-m_1}\alpha_4, \tag{12}$$

这些关系贯穿全文使用。

除了求解方程组（7）之外，二元贝塔分布的参数空间还要求解满足 α_1、……、 $\alpha_4 > 0$。然而，这并非总是能够实现。令 $\mathcal{M} \subseteq [0,1]^3 \times [-1,1]$ 为满足以下条件的值集合（m_1， m_2， v_1， ρ）：方程（9）中的解严格为正，且 $v_1 < m_1(1-m_1)$ ——参见注1。 图1展示了以下形式的区域

$$R_{v_1,\rho} = \{(m_1,m_2) : (m_1,m_2,v_1,\rho) \in \mathcal{M}\},$$

其中，对于每个子图， v_1 和 ρ 均为固定值。

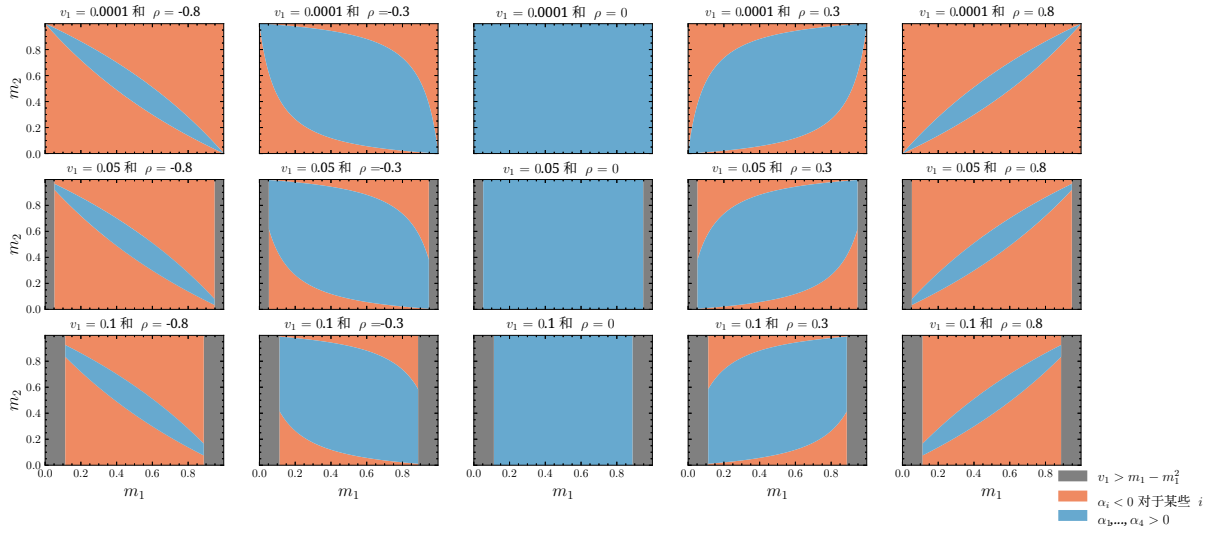


图1：矩方程组的正解：集合 R_{v_1}、 $\rho = \{$（m_1， m_2）的表示：（m_1， m_2， v_1， ρ） $\in \mathcal{M}\}$ 以蓝色表示不同 v_1 和 ρ 值下的区域，其中 $\mathcal{M}$ 是使二元贝塔分布定义良好的值集合（m_1， m_2， v_1， ρ）。橙色表示解（9）中某些 α_i 为负的区域。灰色突出显示方差与均值不满足注1中关系的区域。

根据 注1，我们可以固定 $v_1 < m_1(1-m_1)$，这意味着 $\bar{\alpha} > 0$。将（9）中的解除以 $\bar{\alpha}$，可得 $\alpha_1, \ldots, \alpha_4 > 0$ 当且仅当（iff）以下四个不等式成立时：

$$m_1 m_2 + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)} > 0 \iff \rho > -\frac{m_1 m_2}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}},$$

$$m_1 (1-m_2) - \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)} > 0 \iff \rho < \frac{m_1 (1-m_2)}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}},$$

$$(1-m_1) m_2 - \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)} > 0 \iff \rho < \frac{(1-m_1) m_2}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}},$$

$$(1-m_1)(1-m_2) + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)} > 0 \iff \rho > -\frac{(1-m_1)(1-m_2)}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}},$$

由此得到 ρ的以下区间：

$$\rho \in \left(-\frac{\min(m_1 m_2, (1-m_1)(1-m_2))}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}, \frac{\min(m_1, m_2) - m_1 m_2}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}} \right). \tag{13}$$

例如，当 $m_1 = m_2 = m$ 时，上界为 $\rho_{\max} = 1$，下界为

$$\rho_{\min} = -\frac{\min(m, 1-m)}{\max(m, 1-m)}.$$

O另一方面，如果 $m_1 = 1 - m_2 = m$，下界为 $\rho_{\min} = -1$，上界为 is

$$\rho_{\max} = \frac{\min(m, 1-m)}{\max(m, 1-m)},$$

这与 图1一致。 图 2 展示了不同 m_1 和 m_2取值下的区间长度。 m_1 和 m_2 越远离中心点（0。5，0。5），区间就越短，这揭示了均值与相关系数之间的紧密关系。

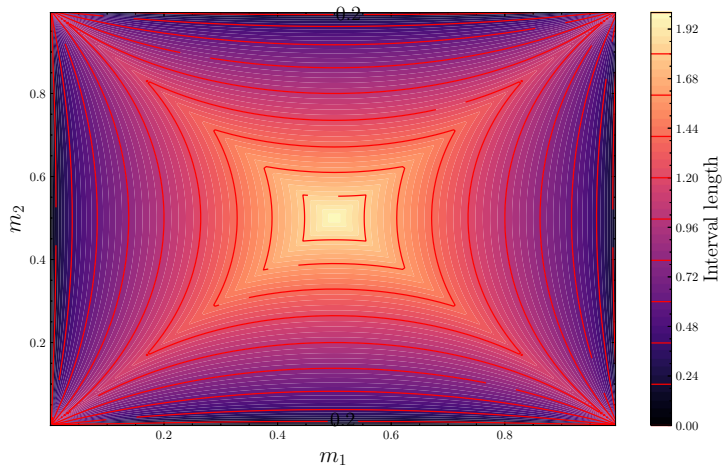


图 2： **ρ的区间长度：** 根据式（ m_1 和 m_2，考虑表达式（13），绘制不同取值下区间长度的等高线图。区间长度越接近 2 越好，这表明在给定均值的情况下，我们有更宽的 ρ 可选范围，从而得到定义良好的双变量贝塔分布。

对命题4 稍作修改，即可解出不含与v_1 和 v_2相关方程的矩方程组，从而得到以下命题：

命题 5（三矩系统）。考虑（7）中去掉 v_1和 v_2相关方程后的方程组，其解为

$$\begin{aligned}
\alpha_1 &= \alpha_4 \frac{m_1 m_2 + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}{(1-m_1)(1-m_2) + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}, \\
\alpha_2 &= \alpha_4 \frac{m_1 (1-m_2) - \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}{(1-m_1)(1-m_2) + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}, \\
\alpha_3 &= \alpha_4 \frac{m_2 (1-m_1) - \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}{(1-m_1)(1-m_2) + \rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)}},
\end{aligned} \tag{14}$$

其中 α_4 是一个自由参数。

该命题的证明使用了关系式（11）和（12），以及符号求解器 SymPy（Meurer 等人，2017），具体代码见A.1 节。此外，我们还对本文中的这一解以及其他所有解进行了数值检验。

3 参数估计

现在，我们将着手制定策略，根据数据 $\boldsymbol{Z} = \{(X_1, Y_1 \ldots, X_n, Y_n)\}$ 来估计双变量贝塔分布的参数向量 $\boldsymbol{\alpha}$，其中 $\boldsymbol{Z} \overset{iid}{\sim}$BivariateBeta($\boldsymbol{\alpha}$)。我们将探讨矩估计法和贝叶斯方法。

3.1 矩估计法

我们首先利用第 2.1 节中给出的结果，采用基于矩估计法的频率学派方法。

记作

$$\begin{aligned}
\bar{X}_n &= \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad P_n = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{(n-1)S_{X,n}S_{Y,n}} \\
S^2_{X,n} &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \quad \text{and} \quad S^2_{Y,n} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2.
\end{aligned} \tag{15}$$

对于这些统计量的观测（样本）版本，我们记为 $\bar{X}_n = \hat{m}_1$、$\bar{Y}_n = \hat{m}_2$、$S^2_{X,n} = \hat{v}_1$、$S^2_{Y,n} = \hat{v}_2$ 和 $P_n = \hat{\rho}$。因此，矩估计法通过匹配理论矩与观测矩来求解 $\boldsymbol{\alpha}$。为此，我们在第 2.1 节的方程中，用（$\bar{X}_n$、$\bar{Y}_n$、$S^2_{X,n}$、$S^2_{Y,n}$、P_n）替换（m_1、m_2、v_1、v_2、ρ）。为了量化估计值的不确定性，我们通过估计标准误，并为 $\boldsymbol{\alpha}$ 的每个分量构造近似的 95% 置信区间，采用一种简单的非参数自助法（Efron，1979），对观测值对进行 $B = 500$ 次重抽样。

四方程解析解的矩估计法

第一种方法直接求解方程组（7），并忽略第二变量的方差，因为否则以概率 1 不会存在解。因此，矩估计法 1（MM1）估计量 $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ 对于 $\boldsymbol{\alpha}$ 为

$$\begin{aligned}
\tilde{\alpha}_4 &= \bar{\alpha}\left(P_n\sqrt{\bar{X}_n\bar{Y}_n(1-\bar{X}_n)(1-\bar{Y}_n)} + (1-\bar{X}_n)(1-\bar{Y}_n)\right) \\
\hat{\alpha}_1 &= \max\{0, (\bar{X}_n + \bar{Y}_n - 1)\bar{\alpha} + \tilde{\alpha}_4\}, \\
\hat{\alpha}_2 &= \max\{0, (1-\bar{Y}_n)\bar{\alpha} - \tilde{\alpha}_4\}, \\
\hat{\alpha}_3 &= \max\{0, (1-\bar{X}_n)\bar{\alpha} - \tilde{\alpha}_4\}, \\
\hat{\alpha}_4 &= \max\{0, \tilde{\alpha}_4\},
\end{aligned} \tag{16}$$

其中

$$\bar{\alpha} = \frac{\bar{X}_n - \bar{X}_n^2 - S^2_{X,n}}{S^2_{X,n}}.$$

由于解可能会根据 $\bar{X}_n$、$\bar{Y}_n$、$S^2_{X,n}$ 和 P_n的取值而为非正，因此在这种情况下，我们将$\hat{\alpha}_i = 0$设为近似值。Arnold 和 Ng（2011）在这些情况下也采用相同的启发式方法。

命题 6（解的符号）。设 $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ 为由观测矩 $\hat{m}_1, \hat{m}_2, \hat{v}_1$ and $\hat{\rho}$ 根据 (9) 得到的解。若 $\hat{v}_1 < \hat{m}_1(1 - \hat{m}_1)$，则至多有一个坐标为非正。

参见附录 A。上述命题意味着 MM1 至多会有一个零值。使用 MM1 从$\boldsymbol{\alpha}$ 估计 $\boldsymbol{Z}$ 后，若其严格为正，我们可以令随机对$\tilde{}$ $\tilde{}(X,\ Y) \sim$服从双变量贝塔分布($\hat{\boldsymbol{\alpha}}$)。于是，

$$\mathrm{Var}(\tilde{X}) = \frac{\mathbb{E}[\tilde{X}](1 - \mathbb{E}[\tilde{X}])}{s_\alpha + 1} \implies s_\alpha + 1 = \frac{\mathbb{E}[\tilde{X}](1 - \mathbb{E}[\tilde{X}])}{\mathrm{Var}(\tilde{X})}.$$

因此，

$$\mathrm{Var}(\tilde{Y}) = \mathrm{Var}(\tilde{X})\frac{\mathbb{E}[\tilde{Y}](1 - \mathbb{E}[\tilde{Y}])}{\mathbb{E}[\tilde{X}](1 - \mathbb{E}[\tilde{X}])},$$

这给出了第二变量方差的公式，因为该估计量没有使用其样本对应值。

三方程解析解的矩估计法

我们考虑的第二个估计量基于命题 5。对于给定的 α_4，注意到 (14) 中的表达式给出

$$s_\alpha = \sum_{j=1}^{4} \alpha_j = \frac{\alpha_4}{(1 - \bar{X}_n)(1 - \bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}.$$

此外，我们希望

$$S^2_{X,n} = \frac{\bar{X}_n(1 - \bar{X}_n)}{s_\alpha + 1}, S^2_{Y,n} = \frac{\bar{Y}_n(1 - \bar{Y}_n)}{s_\alpha + 1} \implies s_\alpha + 1 = \frac{\bar{X}_n(1 - \bar{X}_n)}{S^2_{X,n}} = \frac{\bar{Y}_n(1 - \bar{Y}_n)}{S^2_{Y,n}}. \tag{17}$$

因此，我们将该表达式最小化

$$\left(s_\alpha + 1 - \frac{\bar{X}_n(1 - \bar{X}_n)}{S^2_{X,n}}\right)^2 + \left(s_\alpha + 1 - \frac{\bar{Y}_n(1 - \bar{Y}_n)}{S^2_{Y,n}}\right)^2,$$

其解为

$$\tilde{\alpha_4} = \left((1 - \bar{X}_n)(1 - \bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}\right)\left(\frac{\frac{\bar{X}_n(1-\bar{X}_n)}{S^2_{X,n}} + \frac{\bar{Y}_n(1-\bar{Y}_n)}{S^2_{Y,n}}}{2} - 1\right).$$

那么，矩估计法 2（MM2）估计量 $\hat{\boldsymbol{\alpha}}$ 对于 $\boldsymbol{\alpha}$ 为

$$\begin{aligned}
\hat{\alpha}_1 &= \max\left\{0, \tilde{\alpha}_4\frac{\bar{X}_n\bar{Y}_n + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}{(1 - \bar{X}_n)(1 - \bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}\right\},\\
\hat{\alpha}_2 &= \max\left\{0, \tilde{\alpha}_4\frac{\bar{X}_n(1 - \bar{Y}_n) - P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}{(1 - \bar{X}_n)(1 - \bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}\right\},\\
\hat{\alpha}_3 &= \max\left\{0, \tilde{\alpha}_4\frac{\bar{Y}_n(1 - \bar{X}_n) - P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}{(1 - \bar{X}_n)(1 - \bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1 - \bar{X}_n)(1 - \bar{Y}_n)}}\right\},\\
\hat{\alpha}_4 &= \max\{0, \tilde{\alpha}_4\}.
\end{aligned} \tag{18}$$

注2。 请注意，对于 X_n、Y_n 和 P_n，我们给出了估计量严格为正的充要条件，如*2.1 节*所示。此外，令

$$D = (1-\bar{X}_n)(1-\bar{Y}_n) + P_n\sqrt{\bar{X}_n\bar{Y}_n(1-\bar{X}_n)(1-\bar{Y}_n)}.$$

考虑到*注1*，显然 $\tilde{\alpha} > 0 \iff D > 0$。如果 $P_n > 0$，则有 $D > 0$。于是 $\tilde{\alpha}_4 > 0$并因此 $\alpha_1 > 0$ 且 $\alpha_4 > 0$。此外，如果 $P_n < 0$，那么$\hat{}\alpha_2 > 0$ 且$\hat{}\alpha_3 > 0$，因为 $\tilde{\alpha}_4/D > 0$。因此，分布双变量贝塔分布($\hat{}\boldsymbol{\alpha}$) 可能未定义，因为$\hat{}\alpha_i$ 可能为零，但各阶矩仍有定义。

用解析解求解两个方程的矩估计法

值得探讨的矩估计量的第三种可能性是，利用关系式（11）和（12）来将 $\hat{}\alpha_1$ 和 $\hat{}\alpha_2$ 定义为 $\hat{}\alpha_3$ 和 $\hat{}\alpha_4$ 的函数，依据 X_n 和 Y_n的取值。为求得 $\hat{}\alpha_3$ 和 $\hat{}\alpha_4$的值，我们使用（17）中的表达式以及

$$\rho = \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{\sqrt{(\alpha_1+\alpha_2)(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)(\alpha_3+\alpha_4)}} = \frac{(1-m_1)((\alpha_3+\alpha_4)m_2 - \alpha_3)}{(\alpha_3+\alpha_4)\sqrt{m_1m_2(1-m_1)(1-m_2)}},$$

由此得到表达式

$$\frac{\alpha_3}{\alpha_3+\alpha_4} = m_2 - \frac{\rho\sqrt{m_1m_2(1-m_1)(1-m_2)}}{1-m_1} := E \implies (E-1)\alpha_3 + E\alpha_4 = 0.$$

因此，我们将$\hat{}\alpha_3$ 和$\hat{}\alpha_4$ 定义为该问题的解

$$\begin{aligned} \min_{\alpha_3,\alpha_4>0} &\left(s_\alpha - \frac{\bar{X}_n(1-\bar{X}_n)}{S^2_{X,n}} - 1\right)^2 + \left(s_\alpha - \frac{\bar{Y}_n(1-\bar{Y}_n)}{S^2_{Y,n}} - 1\right)^2 + ((E-1)\alpha_3 + E\alpha_4)^2, \\ \text{s.t. } &(\bar{X}_n + \bar{Y}_n - 1)\alpha_3 + \bar{Y}_n\alpha_4 > 0, \\ &(1-\bar{Y}_n)\alpha_3 + (\bar{X}_n - \bar{Y}_n)\alpha_4 > 0, \end{aligned} \tag{19}$$

其中 $s_\alpha = (\alpha_3+\alpha_4)/(1-X_n)$。这定义了针对 $\boldsymbol{\alpha}$的矩估计法 3 (MM3) 估计量。

无解析解的矩估计法

最后，我们考虑 Olkin 和 Trikalinos (2015) 提出的方法，该方法旨在最小化以下表达式：

$$\begin{aligned} &\left(\bar{X}_n - \frac{\alpha_1+\alpha_2}{s_\alpha}\right)^2 + \left(\bar{Y}_n - \frac{\alpha_1+\alpha_3}{s_\alpha}\right)^2 + \left(P_n - \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{\sqrt{(\alpha_1+\alpha_2)(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)(\alpha_3+\alpha_4)}}\right)^2 + \\ &\left(S^2_{X,n} - \frac{(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)}{s^2_\alpha(s_\alpha+1)}\right)^2 + \left(S^2_{Y,n} - \frac{(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)}{s^2_\alpha(s_\alpha+1)}\right)^2, \end{aligned} \tag{20}$$

满足约束条件 α_1，α_2，α_3，$\alpha_4 > 0$。此外，Olkin 和 Trikalinos (2015) 考虑了额外的限制条件

$$s_\alpha \le \max\left(\frac{\bar{X}_n(1-\bar{X}_n)}{S^2_{X,n}}, \frac{\bar{Y}_n(1-\bar{Y}_n)}{S^2_{Y,n}}\right) - 1, \tag{21}$$

基于表达式 (17)。最小化问题 (20) 的解 $\hat{}\boldsymbol{\alpha}$ 是矩估计法 4 (MM4) 估计量。

注3。 对于每种最小化策略，只要二次损失可微且计算成本低廉，我们就隐式地对其进行了定义。不过，根据具体应用，也可以使用其他损失函数，例如绝对损失、绝对百分比损失或加权二次损失。

简而言之，我们介绍了四种方法：

（MM1）根据命题4求解四个方程；如果解为非正数，则采用启发式方法返回0（Arnold 和 Ng，2011）——至多一个坐标。

（MM2）求解涉及（m_1、 m_2、 ρ）的三个方程，并选择 α_4 ，使相较于v_1 和 v_2的关系最小化。如果解为负数，则采用与MM1相同的方法。

（MM3）求解涉及（m_1、 m_2）的两个方程以及（19）中的优化问题

（MM4）最小化（20）中的表达式，并满足 $\alpha_i > 0$ 和不等式（21）。

3.2 贝叶斯估计

通过为 $\boldsymbol{\alpha}$ 定义先验分布，并使用（3）中的密度，贝叶斯法则通过后验分布根据 $\boldsymbol{Z}$ 更新我们对参数的认知， $p(\boldsymbol{\alpha}|\boldsymbol{Z})$。设定损失函数后，我们可以通过最小化期望后验损失来推导贝叶斯估计量 $\hat{}\boldsymbol{\alpha}$ 。在此，我们将后验均值（二次损失）和中位数（绝对损失）用作点估计策略，分别称为 BE1 和 BE2。密度 $p(\boldsymbol{\alpha}|\boldsymbol{Z})$ 无法解析求得，因此需要借助数值方法计算 EB1 和 EB2。我们采用 Stan 编程语言（由Stan 开发团队， 2022）中实现的动态哈密顿蒙特卡洛（HMC），以获得近似后验样本。

由于数据的联合密度（即似然，方程 3）取决于一个棘手的积分步骤（见 附录 E），因此无法直接应用 HMC 算法。我们通过在模型中加入一个潜变量 $u \in \mathbb{R}^n$ 来绕过这一问题。根据（1）中给出的狄利克雷分布密度，并令（u， x， y） = （u_1， $u_1 + u_2$， $u_1 + u_3$），由变量变换公式可得，

$$f(u, x, y|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} u^{\alpha_1 - 1}(x - u)^{\alpha_2 - 1}(y - u)^{\alpha_3 - 1}(1 - x - y + u)^{\alpha_4 - 1}\mathbb{1}\{u \in \Omega\}.$$

写出完整似然函数 $L(\boldsymbol{\alpha}, u|x, y) = f(x, y|\boldsymbol{\alpha}, u)$，并利用条件密度表达式，可得后验分布

$$\begin{aligned} p(\boldsymbol{\alpha}, u|x, y) &\propto \prod_{i=1}^{n} L(\boldsymbol{\alpha}, u_i|x_i, y_i) f(u_i|\boldsymbol{\alpha})\pi(\boldsymbol{\alpha}), \\ p(\boldsymbol{\alpha}, u|x, y) &\propto \prod_{i=1}^{n} \frac{f(u_i, x_i, y_i|\boldsymbol{\alpha})}{f(u_i|\boldsymbol{\alpha})} f(u_i|\boldsymbol{\alpha})\pi(\boldsymbol{\alpha}), \\ &= \prod_{i=1}^{n} f(u_i, x_i, y_i|\boldsymbol{\alpha})\pi(\boldsymbol{\alpha}). \end{aligned}$$

为了验证计算得到的后验分布，我们采用基于模拟的校准（SBC）方法，该方法基于秩统计量的直方图（Talts 等人， 2018）。在此方法中，我们利用这样一个事实：如果 $\tilde{\boldsymbol{\alpha}} \sim \pi(\boldsymbol{\alpha})$， $\tilde{\boldsymbol{Z}} \sim$双变量贝塔分布（$\tilde{\alpha}$），且 $\{\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^L\} \sim p(\boldsymbol{\alpha}|\tilde{\boldsymbol{Z}})$，则秩统计量

$$r(\{\alpha_i^1, \ldots, \alpha_i^L\}, \tilde{\alpha}_i) = \sum_{l=1}^{L} 1[\alpha_i^l < \tilde{\alpha}_i] \tag{22}$$

在 $[0, L]$ 中的每个 $i = 1$， ……， 4 上均匀分布。采用独立的伽马分布(1， 1)作为先验分布，得到图 3所示结果。这种形式的一个缺点是，HMC 算法内部的辛积分器（蛙跳法）在 $\boldsymbol{\alpha}$ 较小时会出现发散，如图 B.17所示。除了 SBC，其他相关的马尔可夫链蒙特卡洛和 HMC 专用诊断还包括 $\hat{}R$ (Gelman 和 Rubin， 1992）、发散情况、能量（Betancourt， 2016）以及有效样本量（ESS）。所有诊断方法均可在以下网址提供的笔记本中查看：https://github.com/lucasmoschen/bivariate-beta。此外，在数值实验中，我们使用了 2，000 次预热迭代、2，000 次采样迭代，并将适应性 delta 设为 0。9。其他计算设置均采用 Stan 的默认值。

注4。 请注意，给定 u_i 和 $\boldsymbol{\alpha}$ 后，(x_i、 y_i) 与其他值 x_j、 y_j、 u_j （对应 $j \neq i$）相互独立，这使得向量化实现更加高效。尽管参数维度从4 增加到了 $n + 4$，但与在 *Stan* 中每次迭代都通过求积进行积分相比，执行时间缩短了100 倍。

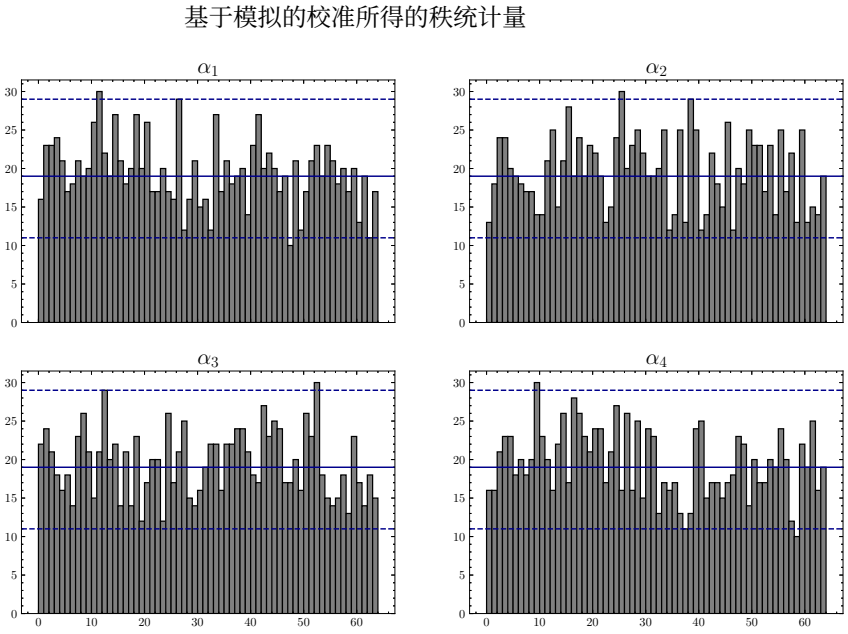


图 3: **基于模拟的校准：** 秩统计量的直方图 r （22）中 $\boldsymbol{\alpha}$各分量的结果。蓝色虚线之间的区域表示离散均匀分布下的 95% 区间，蓝色实线表示中位数。我们将箱数设为 $L = 63$ 并进行$N = 1252$ 次试验。

3.2.1 $\boldsymbol{\alpha}$ 的先验分布

一个自然的问题是，应该为 $\boldsymbol{\alpha}$ 指定什么先验分布。我们考虑两类适当分布族来指定 $\pi(\boldsymbol{\alpha})$，并假设各坐标先验独立 。

(a) $\{\alpha_i\}$ 服从尺度参数为 a_i 、率参数为 b_i 的伽马分布。为简单起见，本文中设定 $a_i = a$ 和 $b_i = b$ 。除非针对每个 α_i 提供不同的信息，否则均采用此设定;

(b) $\{\alpha_i\}$ 的分布密度为

$$\pi(\alpha_i) = \frac{p}{C}\mathbb{1}\{\alpha_i \le C\} + (1-p)\lambda e^{-\lambda(\alpha_i - C)}\mathbb{1}\{\alpha_i > C\}.$$

这是对 $[0, +\infty$ 上均匀分布的适当近似，而后者是非正规的。因此，该先验表示这样一种情形：当 $\alpha_i \in [0, C]$ 时，以概率 p 服从均匀分布；当 $\alpha_i > C$ 时，以概率 $1-p$ 服从率参数为 λ 的指数分布。选择 λ 的值是为了确保密度在 $\alpha_i = C$ 处连续。

我们采用先验预测检验（例如参见 Gabry 等人（2019））来理解每种先验选择所带来的影响。以下是一些有趣的观察结果:

对每个 α_i 使用相同的分布会生成关于相关性的对称分布 ρ，如 图4所示。为了纳入关于 先验符号的信息 ρ，与 α_1 和 α_4 相比，需要为 α_2 和 α_3设定均值更大或更小的伽马分布。此外，为了使 X 和 Y 的均值具有不同的分布， α_2和 α_3 的分布必须不同。

(b) 将均匀-指数分布设为先验时， C 取值过小或过大都不理想。前一种情况会导致双峰分布，而后一种情况则会使 X 和 Y 的方差分布过于集中。此外，改变 C 和 p 的取值，似乎对相关性诱导的先验影响甚微。

更多详情请参见补充材料。

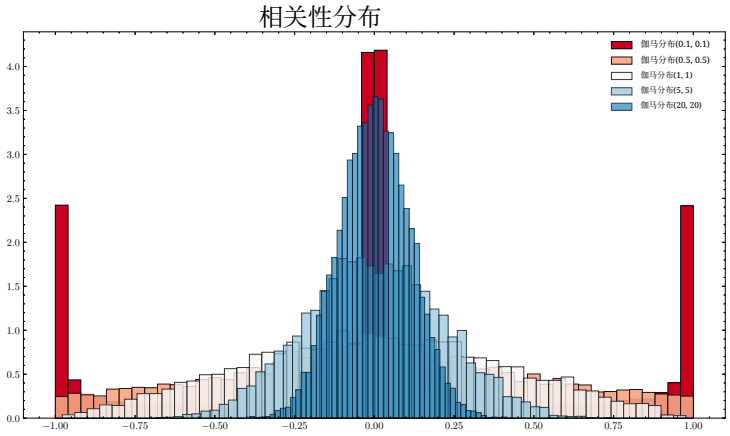


图4：**不同先验分布所诱导的相关性分布：**当 $\alpha_1, \ldots, \alpha_4 \overset{iid}{\sim} \text{Gamma}(a, a)$ 取不同的 a 值时，$\rho = \text{Cor}(X, Y)$ 的分布。

4 双变量贝塔先验的 elicitation

双变量贝塔构造的一个重要应用，是将其用作 $[0,1]^2$ 上变量的先验分布，这些变量被认为或已知先验相关先验地，例如诊断试验的灵敏度和特异度。本节简要讨论如何通过推理其矩和尾部概率等性质，来设定 $\boldsymbol{\alpha}$ 的值，从而确定双变量贝塔分布。

设 $X_1, \ldots, X_m$ 是一个服从分布 P_θ 的随机样本，并假设模型参数 $\theta = (\theta_1, \theta_2)$ 的定义域为 $[0,1]^2$。这里，我们假设 $\theta \sim \text{BivariateBeta}(\boldsymbol{\alpha})$，并针对 $\boldsymbol{\alpha}$ 提出一种 elicitation 方法。为确定参数 $\boldsymbol{\alpha}$，我们考虑研究者希望在 $\boldsymbol{\alpha}$ 的分布中编码以下值：$m_1 = \mathbb{E}[\theta_1]$、$m_2 = \mathbb{E}[\theta_2]$、$v_1 = \text{Var}(\theta_1)$ $v_2 = \text{Var}(\theta_2)$ 以及 $\rho = \text{Cor}(\theta_1, \theta_2)$。根据命题4中的不等式（8），我们已经知道能否用这些矩定义一个二元贝塔分布。因此，我们制定以下策略：

(a) 如果式（9）中的解属于参数空间，我们就将其用作先验参数。

(b) 否则，如果我们认为 v_1 的重要性低于 m_1，m_2 和 ρ，那么只要采用这种方法能得到参数空间中的解，我们就使用 MM2估计量；

(c) 最后，如果上述方法不适用，则应使用 MM3 和 MM4，因为无论输入值如何，它们都能得到参数空间中的解。如果关于 m_1 和m_2 的信息更可靠，则 MM3 更可取；否则，MM4 是一个很有说服力的选择。

尽管我们将关于边际均值和方差的信息定义为输入，但关于边缘分布的每一组信息都可以转换为均值和方差。例如，我们可以给定均值和一个分位数，并将这些信息转换为某个边际变量的均值和方差。

注释 5. 如 第 2 节所述，边缘参数之和等于 s_α，对于 X 和 Y均是如此。对贝塔参数的一种解释是，它们表示实际试验之前的成功次数和失败次数。因此，参数之和就是试验次数或伪试验次数。从这个意义上说，我们对 X 和 Y使用的伪试验次数相同。 ——————

5 数值实验

我们使用 1，000 次蒙特卡洛模拟，通过测量区间估计的偏差、均方误差（MSE）、平均绝对百分比误差（MAPE）、运行时间和覆盖率，比较 $\boldsymbol{\alpha}$ （四种矩估计法和两种贝叶斯估计）这六种估计策略。所有试验均在一台配备英特尔（R）酷睿（TM）i7–1165G7 2.8GHz 处理器（4 核）和 16 GB 内存的 Linux PC 上运行。用于重现试验和使用所提方法的所有计算机代码，均可在 https://github.com/lucasmoschen/bivariate-beta 上以宽松许可证获取。

5.1 双变量贝塔分布下的参数恢复

我们首先考察模型设定正确的情形，即数据来自所研究的双变量贝塔模型。令 $(X, Y) \sim$ BivariateBeta($\boldsymbol{\alpha}$)，其中 $\boldsymbol{\alpha}$ 是未知参数。对于这项试验，我们设定三组具有代表性的参数：$\boldsymbol{\alpha} = (1, 1, 1, 1)$、$\boldsymbol{\alpha} = (2, 7, 3, 1)$ 和$\boldsymbol{\alpha} = (0.7, 0.9, 2.0, 1.5)$。随后，针对样本量为 $n = 50$ 和 $n = 200$个样本的情况，各生成 1,000 次模拟。然后通过蒙特卡洛计算偏差、均方误差（MSE）和平均绝对百分比误差（MAPE）的估计值。对于矩估计法，我们采用非参数自助法进行估计，利用 $B = 500$个自助样本获得 95% 置信区间，随后估计覆盖率。数值结果汇总于附录 C 的表 C.1、C.2 和 C.3 中。

图5和图6显示，MM2、MM3和MM4估计量的平均绝对百分比误差相近，但后者的平均偏差更低。我们还观察到，随着 n 增大，均值误差会减小；这是矩估计法具有相合性的结果，这一性质由大数定律保证，而方程 (9) 中的解是关于 m_1、m_2、v_1 和 ρ的连续函数。MM1的表现不如其他方法，原因有二：它忽略了第二变量方差所包含的信息；并且它精确求解方程，因此容易受到样本矩中噪声的影响，尤其是在 n 不够大时，如图7和图8所示。这些图分别展示了给定 $\boldsymbol{\alpha}$时，$\hat{}\alpha_4$ 和$\hat{}\alpha_3$的估计值与经验矩估计值之间的关系。尤其是，$\hat{}\rho$ 和$\hat{}v_1$ 似乎对估计值的影响最大。此外，尽管我们分别分析每个分量的置信区间，但也可以为整个参数向量构造置信集。图 B.18绘出了自助法得到的 $\boldsymbol{\alpha}$两两估计值。值得注意的是，α_1 和 α_4 的估计值相关，α_2 和 α_3的估计值也是如此。

对于贝叶斯估计，我们使用相互独立的伽马分布 Gamma(1，1) 作为三种设定的先验分布。就平均绝对百分比误差而言，后验均值和中位数表现得更好，但偏差更大，尤其是后验均值；对于贝叶斯估计量而言，这是意料之中的。 图 B.19 展示了当先验分布远离真实值时，贝叶斯估计量的偏差可能有多大。例如，我们表示在观测数据之前，$\mathbb{P}(\alpha_2 > 7) < 0.001$，尽管真实值为 $\alpha_2 = 7$。观测到的置信区间/可信区间覆盖率介于 93

。6% 和 96。3% 之间，接近名义水平 95%。因此，我们得出结论：自助法和后验分布产生了良好的区间估计。需要注意的是，贝叶斯可信区间（BCI）不必具有良好的频率特征，但在本例中确实表现出良好的长期运行特性。然而，当 $\boldsymbol{\alpha} = (2，7、3，1)$ 时，贝叶斯可信区间低估了覆盖率。

最后，图9展示了矩估计法的估计执行时间差异。请注意，当 $\boldsymbol{\alpha} = (2，7，3，1)$ 和 $n = 50$时，MM1 和 MM2 的速度比 MM4 快 50 倍以上，因为它们有显式公式，而 MM4 则取决于初始值与最小值的接近程度。其他设定也呈现出这一规律。

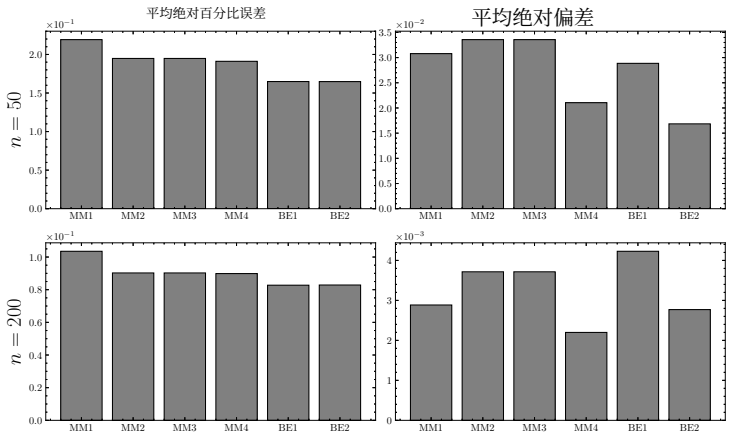


图5：**当真实值为** $\boldsymbol{\alpha} = (1，1，1，1)$ 时的估计性能。平均绝对百分比误差是每个估计值 $\hat{}\alpha_i$ 的平均绝对百分比误差的均值，而平均绝对偏差则是每个估计值偏差绝对值的均值。

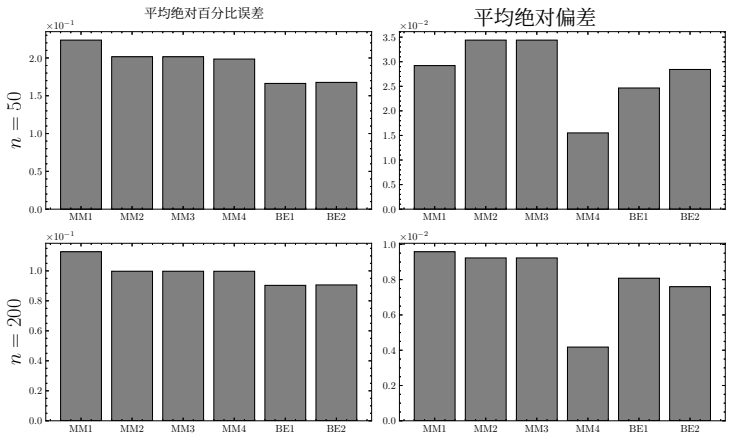


图6：当真实值为 $\boldsymbol{\alpha} = (0.7,0.9,2,1.5)$ 时的**估计性能**。平均绝对百分比误差是每个估计值 $\hat{}\alpha_i$对应的平均绝对百分比误差的均值，而平均绝对偏差则取每个估计值偏差的绝对值的均值。

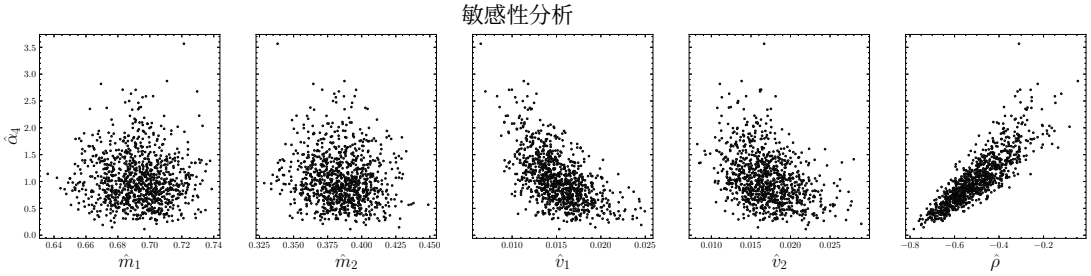


图7：根据 MM1，对大小为 $n = 50$ 和 $\alpha = [2, 7, 3, 1]$ 的不同样本数据集中的 $\hat{}\alpha_4$ **估计值**。我们展示了每个数据集的样本矩与 $\hat{}\alpha_4$估计值之间的散点图。我们观察到，$\hat{}\rho$ 决定了 $\hat{}\alpha_4$的大部分变异性。

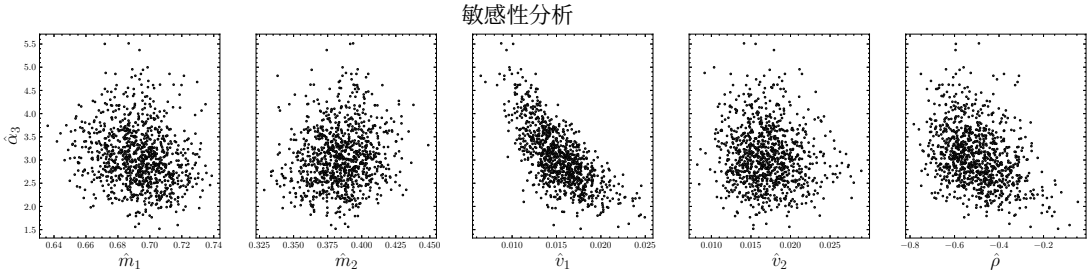


图8：根据 MM1，对大小为 $n = 50$ 和 $\alpha = [2, 7, 3, 1]$ 的不同样本数据集中的 $\hat{}\alpha_3$ **估计值**。我们展示了每个数据集的样本矩与 $\hat{}\alpha_3$估计值之间的散点图。我们观察到，$\hat{}v_1$ 决定了 $\hat{}\alpha_3$的大部分变异性。

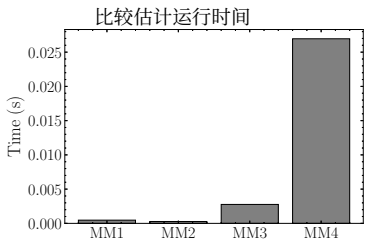


图9：**矩估计法运行时间：**执行矩估计法所需的预计运行时间（以秒计）。

5.2 模型设定错误

现在我们继续研究，当二元模型设定错误时（即真实的数据生成过程是在 $[0,1]^2$ 上的另一种分布），本文提出的方法表现如何。

设 $(X_1,Y_1)\ldots,(X_n,Y_n)$ 为独立同分布的随机变量，其分布的未知密度 $g(x,y)$ 的支撑集覆盖 $[0,1]^2$,，并假设我们使用二元贝塔密度 $f_{\boldsymbol{\alpha}}(x,y)$ 来近似 g。对于矩估计法，我们知道，如果 $\hat{\boldsymbol{\alpha}}$ 的坐标为正，则 (9) 中的解几乎必然收敛到 $\hat{\boldsymbol{\alpha}}$，使得 $f_{\hat{\boldsymbol{\alpha}}}$ 与 g 具有相同的均值、一个边际方差和相关性。不过，仍可能出现一些问题，例如：(i) 方程 (15) 中给出的矩的经验估计量可能具有较大的方差，因而产生较差的估计值；(ii) 根据图1中的结果，即使 n 的值很大，方程组的解也可能为负；(iii) 尽管 g 的矩得到了很好的近似，却可能无法体现其他感兴趣的特征，例如概率。

定义函数 $g(x,y)$ 为 $X=1/(1+e^{-G_1})$ 和 $Y=1/(1+e^{-G_2})$ 的联合密度，使得 $G=(G_1,G_2)$ 且 $G\sim$ 服从正态分布 Normal(μ,Σ)，其中 $\mu\in\mathbb{R}^2$ 和 $\Sigma\in\mathbb{R}^{2\times2}$ 是协方差矩阵 [1]。对于这项试验，我们设定两组具有代表性的参数：$\mu=$ (0, 0) 和 $\Sigma=[[1,0.1],[0.1,1]]$，以及 $\mu=-$(1, 1) 和 $\Sigma=[[2.25,-1.2],[-1.2,1]]$，并模拟 $n=50$ 个样本以获得 $\boldsymbol{\alpha}$ 的估计值。然后，我们通过 1,000 次模拟，使用蒙特卡洛方法估计偏差、平均绝对百分比误差和均方误差，并将其与真实矩和估计矩进行比较，估计方法为 $\hat{\boldsymbol{\alpha}}$。结果汇总于表 C.4 和 C.5 中。在第一个试验中，真实矩约为 $\mathbb{E}[X]=\mathbb{E}[Y]=0.5$，Var$(X)$ $=$Var(Y) $=0.0433$，且 Cor(X,Y) $=0.098$。根据这些数值，由于均值和方差满足命题4给出的关系，MM1 估计值非常精确。其他估计量的表现也非常相似。尽管如此，如图 11 所示，相关系数估计值的平均绝对百分比误差仍大于 100%。这是因为我们用作随机变量间相关系数估计量的样本相关系数，并不是 ρ 的良好估计量。在图 10 中，我们观察到 P_n 的样本分布过于分散。具体而言，我们估计 $\mathbb{P}(P_n\notin[0,0.2])\approx0.5$，也就是说

$$\mathbb{P}\left(\left|\frac{P_n-\rho}{\rho}\right|>1\right)\approx0.5,$$

这解释了相关性较高的原因，因为矩估计法从构造上是对经验相关性而非真实相关性进行近似。

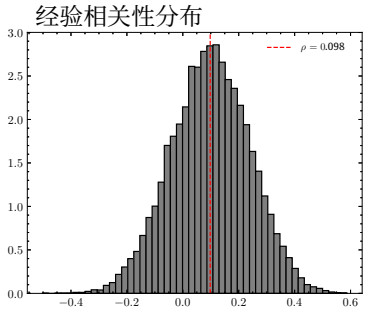


图 10：**相关性的样本分布**：当 $\mu=$(0，0)、$\Sigma=[[1,0.1],[0.1,1]]$ 和 $n=50$ 时，通过蒙特卡罗样本计算得到的经验相关性样本分布P_n。

第二个试验的真实矩为 $\mathbb{E}[X]=0.33$、$\mathbb{E}[Y]=0.3$、Var(X) $=0.062$、Var(Y) $=$0。033 以及 Cor$(X$、$Y)$ $=-0.73$。这些值无法得到定义良好的双变量贝塔分布，因为方程（9）中的解不属于参数空间。因此，每种方法都根据其定义方式提出近似解。在 图 11 中，我们注意到 MM1 在第二个方差上的误差较大，因为它忽略了该方差的值。另一方面，MM4 通过增大均值和第一个方差上的误差进行补偿。在实际应用中，应考虑每种情形下的偏好：例如，如果非常关注精确获得边际均值，而不太关注相关性，那么 MM3 是一个很好的备择假设。通过图 12 所示的偏差，也可以看出这些方法之间的差异。例如，MM3 为均值精确求解方程，因此偏差为零。尽管 MM1 和 MM2 也对这些量求解，但由于解为负值，因此将负值替换为零。对于这些分析，贝叶斯估计与矩估计法的结果相似。

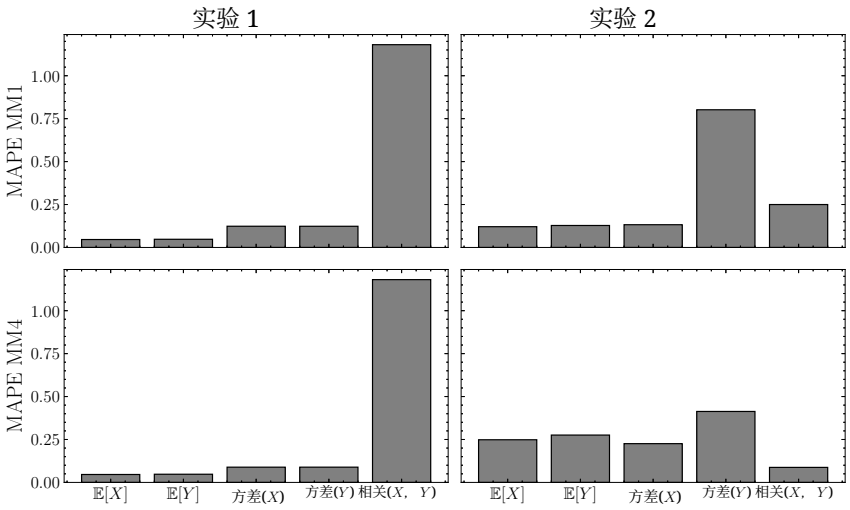


图 11：**模型设定错误情形下各试验的平均绝对百分比误差值**，将真实矩与通过双变量贝塔分布估计得到的矩进行比较，参数 $\hat{\boldsymbol{\alpha}}$ 由 MM1 和 MM4 得出。

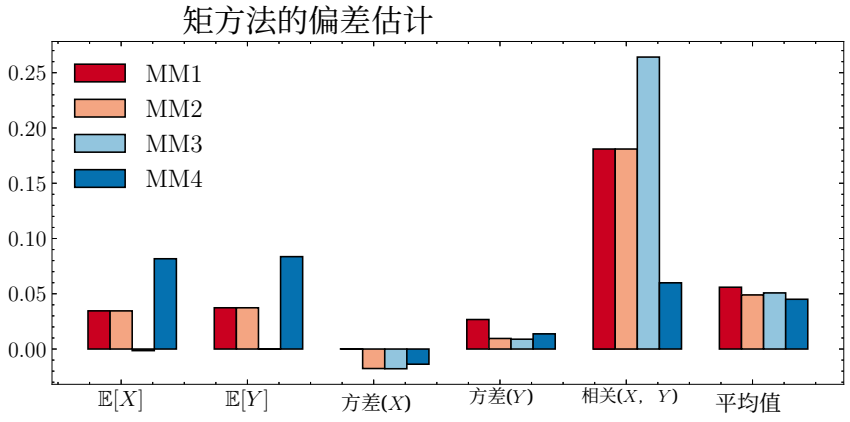


图 12：**各阶矩估计值的偏差：** 对每个矩，我们计算真实值与使用估计参数 $\hat{\boldsymbol{\alpha}}$ 的双变量贝塔分布所对应值之间的偏差，其中 $\mu = [-1, -1]$和 $\Sigma = [[2.25, -1.2], [-1.2, 1]]$。

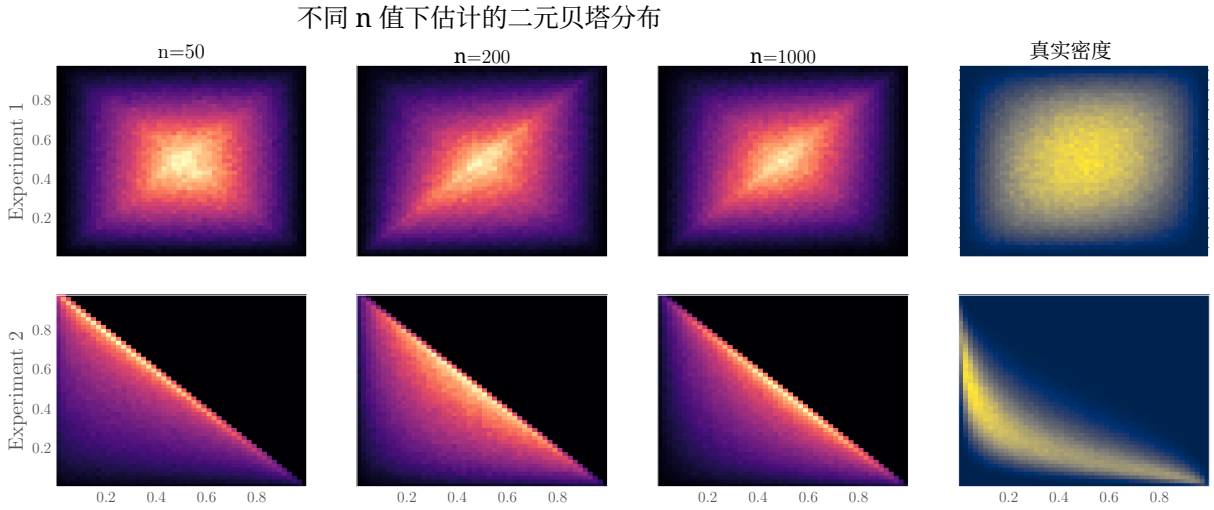


图13：**比较**双变量贝塔密度**的等高线图**，其参数为 $\hat{\boldsymbol{\alpha}}$，并与模型设定错误情形下两项试验的 g 真实密度进行比较。使用 MM4 作为估计量，基于不同数量的样本： $n = 50$，200 和 1，000。

最后，我们将注意力转向想要近似的整体密度——见图13。在第一项试验中，估计矩与真实矩接近，尤其当 n 较大时。两种密度相似，但 g 似乎更分散。比较各分布的边缘峰度后，我们发现，双变量贝塔模型对每个边缘给出的值为 -3，而真实分布的相应值为 -0.85。因此，即使在模型设定正确的情形下，这种分布也可能无法体现其他期望的特征。同样，在第二项试验中，由于矩估计法的估计效果不佳，双变量贝塔分布未能很好地表示该分布的众数。不过，估计分布对真实概率也较低的区域赋予了较小的概率。

5.3 诊断方法

正如我们在第5.2节中所讨论的，二元贝塔分布并不总能很好地近似数据 $\boldsymbol{Z}$的未知分布。这并非分布本身的问题，而是表明它缺乏灵活性。本节提出一种系统的方法，用于诊断给定数据集与二元贝塔分布的兼容性。

5.3.1 渐近诊断

对于第一种诊断方法，我们考虑命题4所建立的关系，

$$\frac{m_1(1-m_1)}{v_1} = \frac{m_2(1-m_2)}{v_2}.$$

这是具有边际均值 m_1 和 m_2 以及边际方差 v_1 和 v_2的二元贝塔分布的必要条件。定义函数

$$g(x_1, x_2, x_3, x_4) = x_1(1-x_1)x_4 - x_2(1-x_3)x_3$$

以及统计量

$$G_n = g(\bar{X}_n, S^2_{X,n}, \bar{Y}_n, S^2_{Y,n}).$$

如果数据来自二元贝塔分布，根据矩估计法的相合性，G_n 当 n 趋于无穷时依概率收敛于 0。此外，正如附录 D 中所证明的，

$$\sqrt{n}G_n \xrightarrow{d} N\left(0, \nabla g(m_1, v_1, m_2, v_2)^T \Sigma \nabla g(m_1, v_1, m_2, v_2)\right).$$

[1]由于 X 和 Y 的矩没有闭式解，我们通过蒙特卡洛方法求得它们。

因此，

$$\sqrt{n}\frac{G_n}{\hat{\sigma}_n} \to N(0,1),$$

其中，$\hat{}\sigma_n^2$ 是 $\nabla g(m_1、\ v_1、\ m_2、\ v_2)^T \Sigma \nabla g(m_1、\ v_1、\ m_2、\ v_2)$的相合估计量。图14给出了使用 $n = 30$ 的数值示例，比较了通过蒙特卡洛估计得到的 $\sqrt{n}G_n/\hat{}\sigma_n$的分布与标准正态分布。该检验程序检验以下原假设：数据来自二元贝塔分布，且 $g(m_1、\ v_1、\ m_2、\ v_2) = 0$；相应的备择假设为 $g(m_1、\ v_1、\ m_2、\ v_2) \neq 0$。如果从大小为 n 的随机样本中观测到 $\sqrt{n}G_n/\hat{}\sigma_n = s$，则 p 值为 $p = 2\Phi(-|s|)$，其中 Φ 是标准正态累积分布函数（CDF）。

图B.20 比较了在原假设和备择假设下的 p值分布：以二元贝塔分布为控制对象，并采用四种不同的独立贝塔设定，以观察统计量的表现。

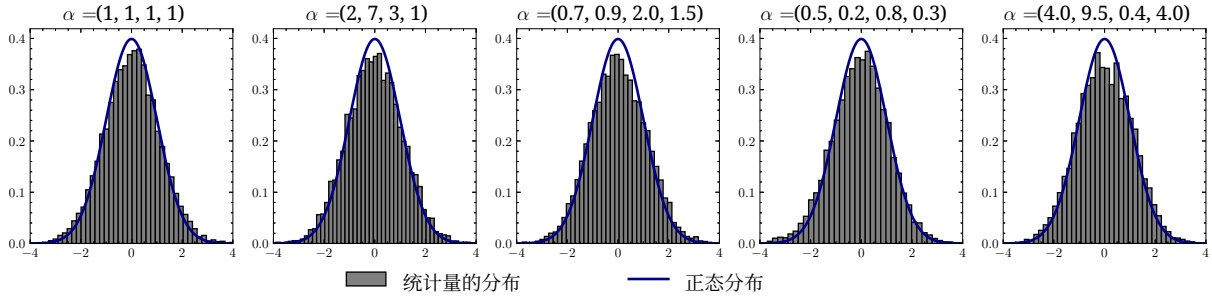


图14：**统计量 ˆ 的概率密度函数与**在 $n = 30$ 条件下，参数 $\boldsymbol{\alpha}$ 的五种设定所对应的标准正态分布的比较。

我们介绍的第二种诊断方法捕捉了命题4中的相关性方面。如果数据来自二元贝塔分布，其边际均值、方差和相关性必须满足式（9）中的关系，且该关系存在正解。设 $S^2_{X,\ n} < \bar{X}_n(1-\bar{X}_n)$，并令 $\hat{}\boldsymbol{\alpha}$ 为方程组 ˆ 在式（9）中的解。定义 $\beta_i = \hat{}\alpha_i/\bar{\alpha}$ 以及汇总统计量

$$M = \min\{\hat{\beta}_1, \ldots, \hat{\beta}_4\}.$$

由于 $M > 0 \iff \hat{}\alpha_i > 0,\ i = 1,\ \ldots\ldots,\ 4$，若对某个固定的 c，$M \leq c$，则所提出的方法拒绝“数据来自二元贝塔分布”这一原假设。我们注意到，当 n 趋于无穷时，考虑到 $c = 0$，该检验不会错误地拒绝原假设，因为矩估计法具有一致性。对于有限的 n，如果某些 α_i 较小，矩估计法可能会得到负解。

分布 M 很难推导，即使在渐近意义下也是如此，因为它是相关统计量的最小值。因此，采用非参数自助法是很有吸引力的。图 B.21 显示了分布 M 在 $n = 50$ 取自 $\boldsymbol{\alpha}$ 中 $[0, 0.5]^4$的均匀随机值时，其分布的第5分位数。这个小值区域是矩估计法产生负值的主要原因。根据经验，我们注意到，如果设定 $c = -0.05$，则在原假设下的拒绝概率至多为 0。05。总之，如果 $M \leq -0.05$，我们将拒绝数据来自二元贝塔分布这一假设。图 B.21 还显示了第1分位数和第10分位数的分布，但二者差异不大。

5.3.2 贝叶斯诊断

诊断贝叶斯估计量时，我们可以使用一些工具。后验预测检验（PPC）利用参数的后验样本生成变量，并将其与样本中的观测值进行比较。在本例中，我们可以将观测矩 $\hat{}m_1, \hat{}m_2, \hat{}v_1, \hat{}v_2$ 和 $\hat{}\rho$ 与通过后验分布的样本计算得到的值进行比较 $p(\boldsymbol{\alpha}|\boldsymbol{Z})$。图15 展示了一个后验预测检验的示例。我们注意到，观测矩位于第2.5和第97.5四分位数之间的区间内。先验信息可以改善这些结果。

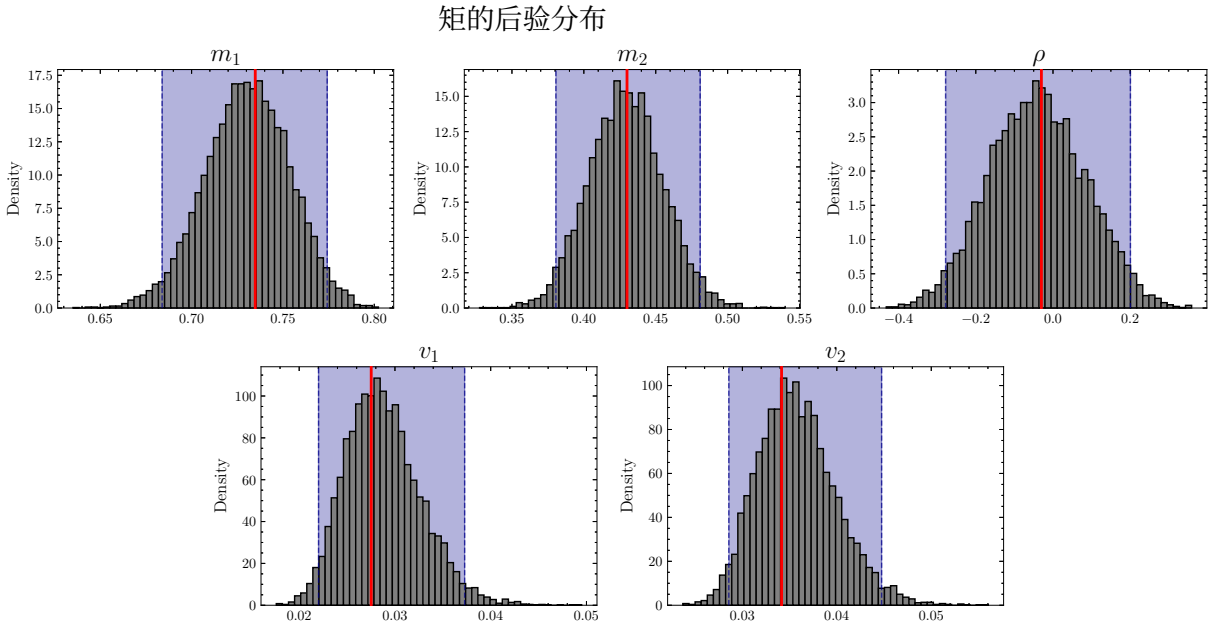


图15：**后验预测检验：**比较矩的后验分布（灰色直方图）与观测值（红色实线）。我们设定 $\boldsymbol{\alpha} = (2,3,0.7,1)$，并使用 $n = 50$ 生成数据，先验分布为相互独立的伽马分布(1,1)。蓝色区域为95%等尾概率区间。

6 讨论与结论

本文探讨了二元分布的一种特定构造方式（Olkin 和 Trikalinos，2015），在参数丰富度与可处理性之间取得了平衡。我们利用对矩的详细数学分析，提出了一类用于估计感兴趣参数 $\boldsymbol{\alpha} \in \mathbb{R}^4_+$的基于矩的估计量。我们还利用潜变量构造，提出了后验分布的一种高效表示方式，以获得贝叶斯估计。最后，我们讨论了如何构造简单而有效的诊断方法，从频率学派和贝叶斯两种视角评估模型拟合。

我们发现，没有一种方法在所有情况下都表现最佳，但统计性能与计算性能之间存在固有的权衡：基于数值求解较少方程的矩估计量通常运行时间较短，但在偏差或平均绝对百分比误差方面表现也较差。总体而言，双变量贝塔分布提供了一种简单的模型，可以快速拟合数据；但我们在模型设定错误情形下的试验表明，它有时可能是一种相当不灵活的分布，无法正确捕捉数据坐标之间的相关性，而这正是任何二元模型的重要特征。

未来研究将着重于检验大量二元模型，包括本文分析的真实数据模型。我们特别关注灵敏度/特异度数据，这类数据在贝叶斯范式下分析不完美检测数据时很有用（Gelman 等人，2020）。

总之，我们希望为应用研究人员提供一套工具，使他们能够快速将双变量贝塔模型拟合到数据并诊断问题；同时，也希望向方法学界介绍一套现代化、完整的工具，用于指导相关比例情形下的模型开发与评估。

致谢

感谢 Rodrigo Targino 和 Eduardo Mendes 富有见地的讨论。第一作者感谢应用数学学院（FGV EMAp）提供的经费支持。

参考文献

Arellano-Valle, R. B., Harnik, S. B., and Genton, M. G. (2021). 多元样本矩的渐近联合分布。载于 Advances in Statistics-Theory and Applications，第 181–206 页。Springer. Arnold, B. C. and Ng, H. K. T. (2011). 灵活的双变量贝塔分布。Journal of Multivariate Analysis, 102(8):1194–1202. Balakrishnan, N. and Lai, C. D. (2009). 连续双变量分布。Springer Science & Business Media. Barros, O. A. d. (2015). Estima ̧cao dos par ̂metros da distribui ̧cao beta bivariada: aplica ̧coes em severidade de doen ̧cas em plantas. 博士论文，Universidade de S ̃ao Paulo. Betancourt, M. (2016). 诊断哈密顿蒙特卡洛中的次优余切分解。arXiv 预印本 arXiv:1604.00695. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan：一种概率编程语言。Journal of statistical software, 76(1). Crackel, R. and Flegal, J. (2017). 灵活的一类双变量贝塔分布的贝叶斯推断。Journal of Statistical Computation and Simulation, 87(2):295–312. Dahabreh, I. J., Trikalinos, T. A., Lau, J., and Schmid, C. (2013). 对检测准确性荟萃分析中二元方法的实证评估。Rockville (MD): Agency for Healthcare Research and Quality (US). Efron, B. (1979). 自助法：重新审视刀切法。The Annals of Statistics, 7(1):1–26. Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). 贝叶斯工作流中的可视化。Journal of the Royal Statistical Society: Series A (Statistics in Society). Gelman, A., Carpenter, B., et al. (2020). 对特异度和灵敏度未知的检测进行贝叶斯分析。Journal of the Royal Statistical Society Series C, 69(5):1269–1283. Gelman, A. and Rubin, D. B. (1992). 使用多条序列从迭代模拟中进行推断。Statistical science, pages 457–472. Libby, D. L. and Novick, M. R. (1982). 用于效用评估的多元广义贝塔分布。Journal of Educational Statistics, 7(4):271–294. Magnussen, S. (2004). 一种生成边缘分布已知且相关性指定的正相关贝塔分布随机变量的算法。Computational statistics & data analysis, 46(2):397–406. ˇ Meurer, A., Smith, C. P., Paprocki, M., Cert ́ık, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J. K., Singh, S., Rathnayake, T., Vig, S., Granger, B. E., Muller, R. P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M. J., Terrel, A. R., Rou ̆cka, v., Saboo, A., Fernando, I., Kulal, S., Cimrman, R., and Scopatz, A. (2017). Sympy：Python 符号计算。PeerJ Computer Science, 3:e103. Nadarajah, S. et al. (2007). 一种新的双变量贝塔分布及其在干旱数据中的应用。Metron-International Journal of Statistics, 65(2):153–174. Nadarajah, S. and Kotz, S. (2005). 一些双变量贝塔分布。Statistics, 39(5):457–466. Nadarajah, S., Shih, S. H., and Nagar, D. K. (2017). 一种新的双变量贝塔分布。Statistics, 51(2):455–474. 出版商：Taylor & Francis 电子版：https://doi.org/10.1080/02331888.2016.1240681. Ng, K. W., Tian, G.-L., and Tang, M.-L. (2011). 狄利克雷分布及相关分布：理论、方法与应用。John Wiley & Sons.

Olkin, I. 和 Liu, R.（2003）。一种二元贝塔分布。《统计与概率通讯》，62(4)：407–412。

Olkin, I. 和 Trikalinos, T. A.（2015）。一种二元贝塔分布的构造。《统计与概率通讯》，96：54–60。

Sarabia, J. M. 和 Castillo, E.（2006）。基于广义三参数贝塔分布的二元分布。载于《分布理论进展、次序统计量与推断》，第85–110页。Springer。

Stan 开发团队（2022）。Stan 建模语言用户指南与参考手册，2.29。 https://mc-stan.org。

Talts, S.、Betancourt, M.、Simpson, D.、Vehtari, A. 和 Gelman, A.（2018）。通过基于模拟的校准验证贝叶斯推断算法。arXiv 预印本 arXiv:1804.06788。

Trick, S., J¨akel, F., and Rothkopf, C. A. (2021). 一种具有任意贝塔边际分布的二元贝塔分布及其向相关狄利克雷分布的推广。arXiv 预印本 arXiv:2104.08069。

A 证明

命题 1 (二元贝塔密度)。 相对于 $\mathbb{R}^2$ 上的勒贝格测度，(X, Y) 的联合密度为

$$f_{X,Y}(x,y) = \frac{1}{B(\boldsymbol{\alpha})} \int_{\Omega} u^{\alpha_1-1}(x-u)^{\alpha_2-1}(y-u)^{\alpha_3-1}(1-x-y+u)^{\alpha_4-1}\, du, \tag{3}$$

其中

$$\Omega = (\max(0, x+y-1), \min(x,y)).$$

命题 命题 *1*的证明。 注意

$$\begin{bmatrix} U_1 \\ X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}.$$

根据变量变换公式,

$$\begin{aligned} f_{U_1,X,Y}(u_1,x,y) &= f_{U_1,U_2,U_3}(u_1, x-u_1, y-u_1), \\ &= \frac{1}{B(\boldsymbol{\alpha})} u_1^{\alpha_1-1}(x-u_1)^{\alpha_2-1}(y-u_1)^{\alpha_3-1}(1-x-y+u_1)^{\alpha_4-1}, \end{aligned} \tag{23}$$

其中 $0 \le u_1 \le x$， $u_1 \le y$，以及 $0 \le 1-x-y+u_1$。因此,

$$f_{X,Y}(x,y) = \frac{1}{B(\boldsymbol{\alpha})} \int_{\Omega} u_1^{\alpha_1-1}(x-u_1)^{\alpha_2-1}(y-u_1)^{\alpha_3-1}(1-x-y+u_1)^{\alpha_4-1}\, du_1, \tag{24}$$

使得 $\Omega = \{u_1 \in [0,1] : \max(0,\ x+y-1) < u_1 < \min(x,\ y)\}$。 □

命题4 (矩系统的解)。 方程组 (7) 无解，当

$$\frac{m_1(1-m_1)}{v_1} \neq \frac{m_2(1-m_2)}{v_2}. \tag{8}$$

此外 ，去掉关于 v_2 的第四个方程后，该方程组有唯一解，给定为 by

$$\begin{aligned} \alpha_1 &= (m_1+m_2-1)\bar{\alpha} + \alpha_4, \\ \alpha_2 &= (1-m_2)\bar{\alpha} - \alpha_4, \\ \alpha_3 &= (1-m_1)\bar{\alpha} - \alpha_4, \\ \alpha_4 &= \bar{\alpha}\left(\rho\sqrt{m_1 m_2 (1-m_1)(1-m_2)} + (1-m_1)(1-m_2)\right), \end{aligned} \tag{9}$$

其中

$$\bar{\alpha} = \frac{m_1 - m_1^2 - v_1}{v_1} = s_\alpha. \tag{10}$$

命题*4*的证明命题*4*。 注意，由于以下原因，我们可以简化第三个和第四个方程

$$\frac{\alpha_3+\alpha_4}{s_\alpha}=\frac{s_\alpha-(\alpha_1+\alpha_2)}{s_\alpha}=1-m_1,$$

类似地,

$$\frac{\alpha_2+\alpha_4}{s_\alpha}=1-m_2.$$

因此,

$$v_1=\frac{m_1(1-m_1)}{s_\alpha+1},$$
$$v_2=\frac{m_2(1-m_2)}{s_\alpha+1}.$$

这已经说明，如果.....，方程组就没有解

$$\frac{m_1(1-m_1)}{v_1}\neq\frac{m_2(1-m_2)}{v_2}.$$

方程组（7）的前两个方程可以改写为一个线性方程组:

$$(m_1-1)\alpha_1+(m_1-1)\alpha_2+m_1\alpha_3+m_1\alpha_4=0,$$
$$(m_2-1)\alpha_1+m_2\alpha_2+(m_2-1)\alpha_3+m_2\alpha_4=0,$$

这等价于

$$\alpha_1+\alpha_2+\frac{m_1}{m_1-1}\alpha_3+\frac{m_1}{m_1-1}\alpha_4=0,$$
$$\alpha_2+\frac{1-m_2}{m_1-1}\alpha_3+\frac{m_1-m_2}{m_1-1}\alpha_4=0.$$

然后，我们可以将 α_1 和 α_2 表示为关于 α_3 和 α_4 的函数:

$$\alpha_1=\frac{m_1+m_2-1}{1-m_1}\alpha_3+\frac{m_2}{1-m_1}\alpha_4$$
$$\alpha_2=\frac{1-m_2}{1-m_1}\alpha_3+\frac{m_1-m_2}{1-m_1}\alpha_4.$$

根据上面的表达式，记 $\alpha_1=a_3\alpha_3+a_4\alpha_4$、 $\alpha_2=b_3\alpha_3+b_4\alpha_4$、 $c_3=a_3+b_3+1$和$c_4=a_4+b_4+1$。则第三个方程可改写为

$$s_\alpha=a_3\alpha_3+a_4\alpha_4+b_3\alpha_3+b_4\alpha_4+\alpha_3+\alpha_4=c_3\alpha_3+c_4\alpha_4=\frac{m_1(1-m_1)}{v_1}-1,$$

这意味着

$$\alpha_3=\frac{m_1(1-m_1)-v_1-c_4v_1\alpha_4}{c_3v_1},$$

这是关于 α_4的线性函数。经过一些简化后，我们将关于 α_4 的函数表达式总结如下:

$$\alpha_1=(m_1+m_2-1)\frac{(m_1-m_1^2-v_1)}{v_1}+\alpha_4,$$
$$\alpha_2=(1-m_2)\frac{(m_1-m_1^2-v_1)}{v_1}-\alpha_4,$$
$$\alpha_3=(1-m_1)\frac{(m_1-m_1^2-v_1)}{v_1}-\alpha_4,$$

然后,

$$s_\alpha=\frac{m_1-m_1^2-v_1}{v_1}.$$

现在利用方程组（7）中的前两个方程和上面的方程，将第五个方程改写如下

$$\begin{aligned}\rho &= \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{\sqrt{(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)}},\\ &= \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{s_\alpha^2\sqrt{m_1m_2(1-m_1)(1-m_2)}},\\ &= \frac{(m_1+m_2-1)s_\alpha\alpha_4 + \alpha_4^2 - ((1-m_2)s_\alpha - \alpha_4)((1-m_1)s_\alpha - \alpha_4)}{s_\alpha^2\sqrt{m_1m_2(1-m_1)(1-m_2)}},\\ &= \frac{\alpha_4 - (1-m_1)(1-m_2)s_\alpha}{s_\alpha\sqrt{m_1m_2(1-m_1)(1-m_2)}},\end{aligned} \tag{25}$$

因此，解为

$$\alpha_4 = s_\alpha\left(\rho\sqrt{m_1m_2(1-m_1)(1-m_2)} + (1-m_1)(1-m_2)\right),$$

证毕。 □

命题 6（解的符号）。设 $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ 是由（9）及观测矩 $\hat{m}_1$、$\hat{m}_2$、$\hat{v}_1$ 和 $\hat{\rho}$ 得到的解。若 $\hat{v}_1 < \hat{m}_1(1-\hat{m}_1)$，则至多有一个坐标是非正的。

ˆ命题 6的证明。由于 $\bar{\alpha} > 0$ 根据假设，定义 $b_i = \hat{\alpha}_i/\bar{\alpha}$，然后注意到以下关系式成立，这是由矩估计解的定义直接得到的:

1. ˆ $\beta_1 + \hat{\beta}_2 = \hat{\alpha}_1 + \hat{}_{_}^{\alpha_2} = \bar{X}_n > 0;\alpha$
2. ˆ $\beta_1 + \hat{\beta}_3 = \hat{\alpha}_1 + \hat{}_{_}^{\alpha_3} = \bar{Y}_n > 0;\alpha$
3. ˆ $\beta_2 + \hat{\beta}_4 = 1 - \hat{\alpha}_1 + \hat{}_{_}^{\alpha_3} = 1 - \bar{Y}_n > 0;\alpha$4. ˆ $\beta_3 + \hat{\beta}_4 = 1 - \hat{\alpha}_2 + \hat{}_{_}^{\alpha_3} = 1 - \bar{X}_n > 0;\alpha$

5. ˆ $\beta_1 + \hat{\beta}_4 = \hat{\alpha}_1 + \hat{}_{\bar{\alpha}}^{\alpha_4} = 2P_n\sqrt{\bar{X}_n\bar{Y}_n(1-\bar{X}_n)(1-\bar{Y}_n)} + \bar{X}_n\hat{m}_2 + (1-\bar{X}_n)(1-\bar{Y}_n) > 0$，考虑到 $P_n > -1$ 并将剩余表达式最小化。此外，考虑到 $P_n < 1$，我们 ˆ ˆ ˆ ˆ 得出 $\beta_1 + \beta_4 < 1$，这意味着 $\beta_2 + \beta_3 > 0$。

我们可以 ˆ ˆ 因此可以得出结论：若对于某个 j，$\beta_j \le 0$，则对于 $i \ne j$，$\beta_i > 0$。换言之，至多有一个 e 坐标 是非正的。 □

A.1 三方程组的解的代码

```
import sympy as sp

if __name__ == '__main__':m1, m2, rho, alpha3, alpha4 = sp.symbols('m1 m2 rho alpha3 alpha4')
alpha1 = (m1 + m2 - 1)/(1 - m1) * alpha3 + m2/(1 - m1) * alpha4
alpha2 = (1 - m2)/(1 - m1) * alpha3 + (m1 - m2)/(1 - m1) * alpha4
alpha_sum = sp.simplify(alpha1 + alpha2 + alpha3 + alpha4)
expression = rho - (alpha1 * alpha4 - alpha2 * alpha3)
expression /= (alpha_sum**2 * sp.sqrt(m1 * m2 * (1-m1) * (1 - m2)))
alpha3 = sp.simplify(sp.solve(expression, alpha3)[0])
alpha1 = sp.simplify((m1 + m2 - 1)/(1 - m1) * alpha3 + m2/(1 - m1) * alpha4)
alpha2 = sp.simplify((1 - m2)/(1 - m1) * alpha3 + (m1 - m2)/(1 - m1) * alpha4)
print('alpha1 = {}'.format(alpha1))print('alpha2 = {}'.format(alpha2))
print('alpha3 = {}'.format(alpha3))
```

B 附加图

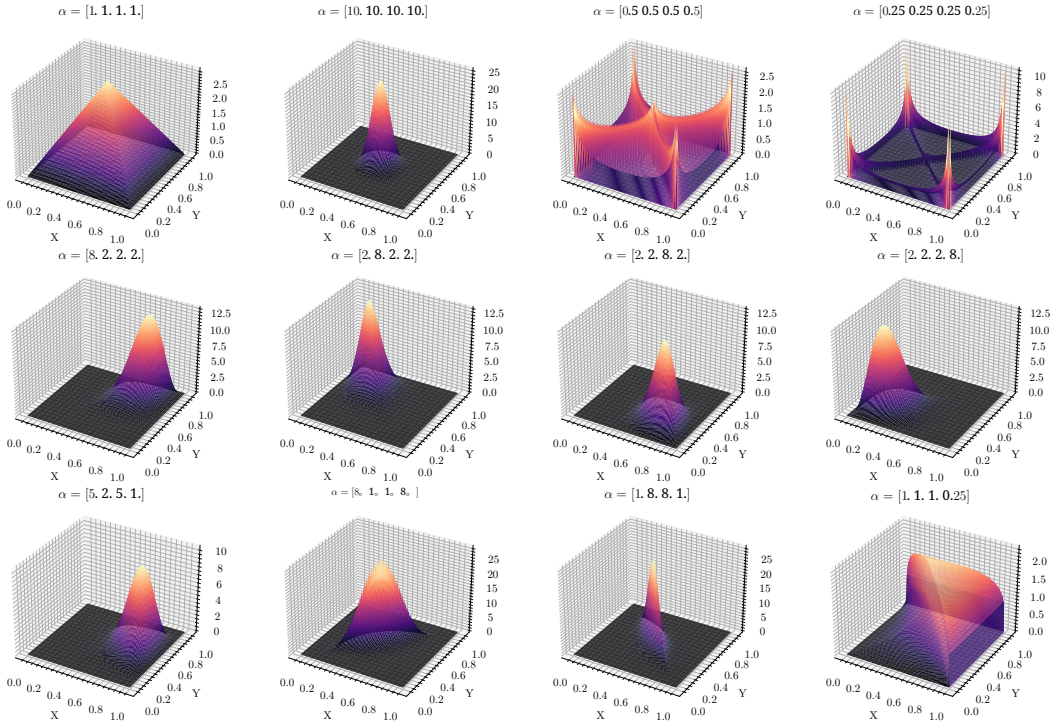


图 B.16：**变量** X **和** Y的联合密度：不同 $\boldsymbol{\alpha}$ 取值下的二元密度。第一行的四幅图关于众数对称，且变量 X 和 Y之间没有相关性。

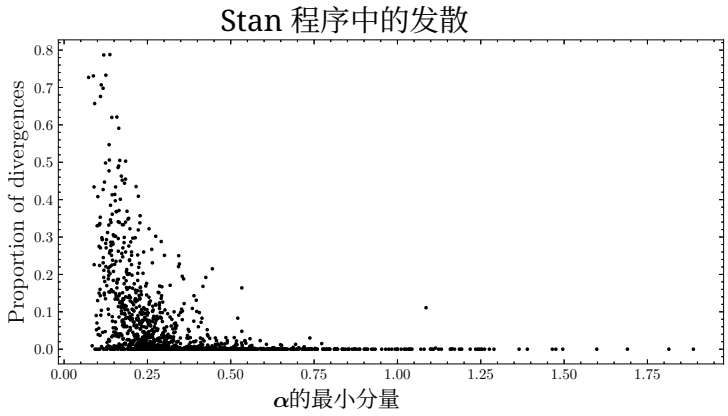


图 B.17：**Stan 程序中的发散：**基于模拟的校准中每个抽样 $\tilde{\boldsymbol{\alpha}}$ 的 Stan 程序发散比例。如果某些 $\alpha_i < 0.5$ 的发散次数超出预期，就会导致结果产生偏差。

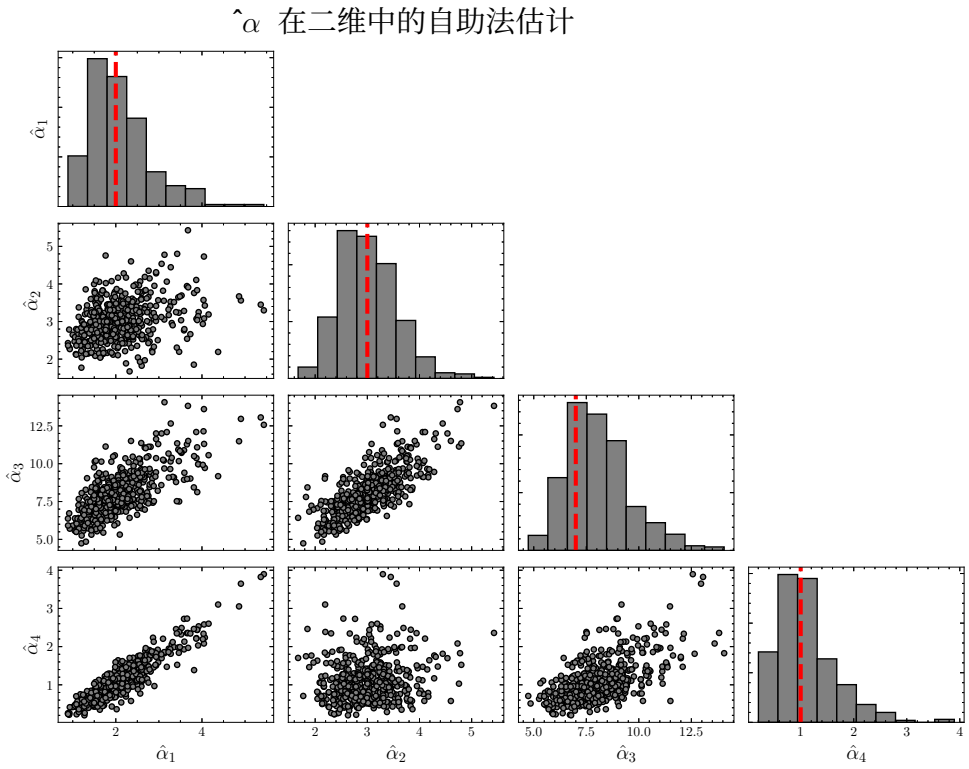


图 B.18：**二维中的自助法估计：**使用带有 $B = 500$ 的自助法得到的 $\boldsymbol{\alpha}$ 各分量的成对估计值。红线表示生成大小为 $n = 50$ 的原始数据集的 $\boldsymbol{\alpha} =$ 真实值（2，3，7，1）。

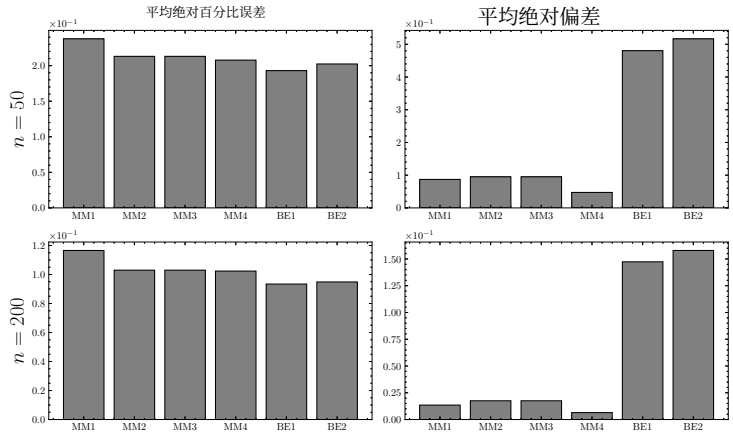
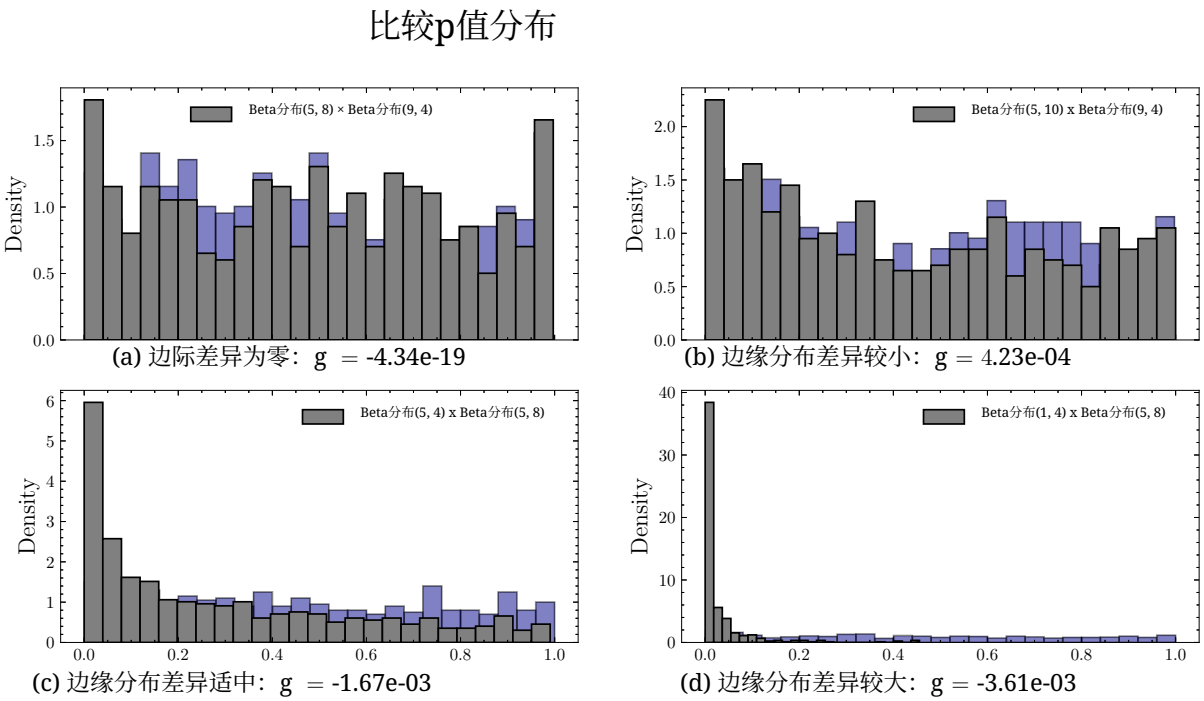


图 B.19：**估计性能** 当真实值为 $\boldsymbol{\alpha} = (2,7,3,1)$时。平均绝对百分比误差是各估计值$\hat{\alpha}_i$对应的平均绝对百分比误差的均值，而平均绝对偏差则是各估计值偏差绝对值的均值。



图B.20：**比较p值分布：** 我们比较了在“数据来自参数为$(2,3,7,1)$的二元贝塔分布”这一假设下的p值分布（蓝色），以及与指定参数对应的两个独立贝塔分布的p值分布。

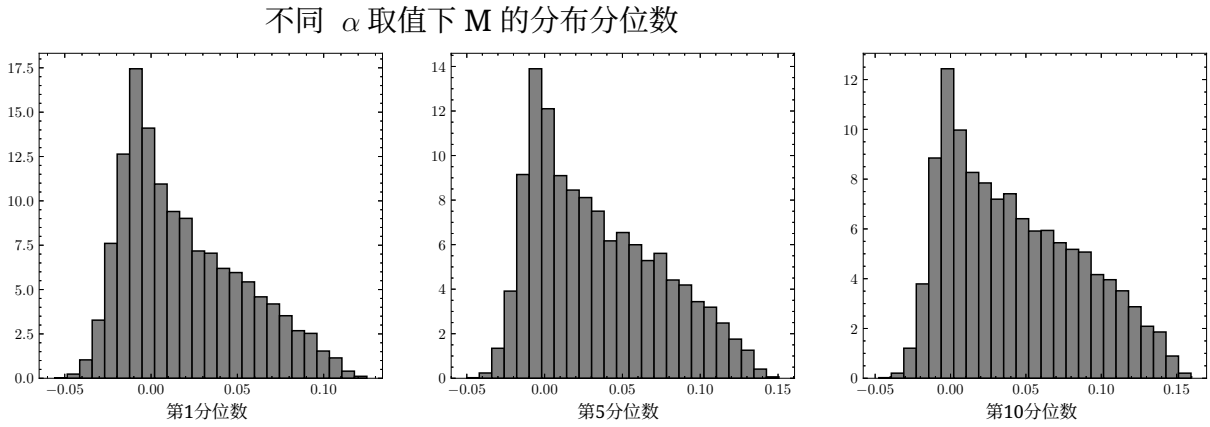


图 B.21：**分位数分布：** $[0, 0.5]^4$ 中均匀随机选取的 $\boldsymbol{\alpha}$ 不同设定下，统计量 M 的分布的第1、第5和第10分位数。

C 附加表格

		$n=50$				$n=200$			
方法	评估	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4
MM1	偏差 (10^{-2})	1.79	4.08	4.01	2.43	0.23	0.33	-0.29	0.31
	均方误差 (10^{-1})	0.78	0.78	0.76	0.84	0.17	0.17	0.17	0.18
	平均绝对百分比误差 (10^{-1})	2.23	2.18	2.11	2.25	1.04	1.05	1.01	1.04
	覆盖率（%）	94.7	93.9	95.1	93.0	94.4	93.5	94.1	93.2
MM2	偏差 (10^{-2})	2.12	4.32	4.31	2.68	0.41	0.52	-0.09	0.46
	均方误差 (10^{-1})	0.63	0.6	0.59	0.66	0.13	0.13	0.13	0.13
	平均绝对百分比误差 (10^{-1})	2.01	1.91	1.87	2.02	0.9	0.91	0.9	0.9
	覆盖率（%）	93.0	93.8	94.3	92.4	94.4	94.3	95.3	95.5
MM3	偏差 (10^{-2})	2.12	4.32	4.31	2.68	0.41	0.52	-0.09	0.46
	均方误差 (10^{-1})	0.63	0.6	0.59	0.66	0.13	0.13	0.13	0.13
	平均绝对百分比误差 (10^{-1})	2.01	1.91	1.87	2.02	0.9	0.91	0.9	0.9
	覆盖率（%）	93.8	94.0	94.4	92.8	95.2	94.4	94.7	95.5
MM4	偏差 (10^{-2})	0.88	3.06	3.04	1.44	0.11	0.22	-0.39	0.16
	均方误差 (10^{-1})	0.61	0.57	0.56	0.63	0.13	0.13	0.13	0.13
	平均绝对百分比误差 (10^{-1})	1.98	1.87	1.82	1.98	0.89	0.91	0.9	0.9
	覆盖率（%）	93.3	93.8	94.8	93.1	95.1	95.1	95.1	95.2
BE1	偏差 (10^{-2})	1.91	3.71	3.65	2.27	0.53	0.44	-0.14	0.59
	均方误差 (10^{-1})	0.45	0.44	0.42	0.46	0.11	0.11	0.11	0.11
	平均绝对百分比误差 (10^{-1})	1.67	1.65	1.58	1.7	0.81	0.84	0.83	0.83
	覆盖率（%）	94.6	95.9	96.1	95.2	95.3	94.4	94.8	95.5
BE2	偏差 (10^{-2})	0.7	2.51	2.45	1.07	0.24	0.14	-0.43	0.3
	均方误差 (10^{-1})	0.44	0.44	0.41	0.46	0.11	0.11	0.11	0.11
	平均绝对百分比误差 (10^{-1})	1.67	1.64	1.58	1.7	0.82	0.84	0.83	0.83
	覆盖率（%）	94.6	95.9	96.1	95.2	95.3	94.4	94.8	95.5

表 C.1：当 α 生成过程的真实值为 $\alpha=(1,\ 1,\ 1,\ 1)$，样本数为 $n=50$ 或 $n=200$时，六种方法各自的偏差估计、均方误差、平均绝对百分比误差和覆盖率。估计值采用蒙特卡洛方法计算，共迭代 1，000 次，具体如第 5.1 节所述。

		$n=50$				$n=200$			
方法	评估	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4
MM1	偏差 (10^{-2})	6.44	17.56	6.58	4.26	1.48	1.81	0.61	1.5
	均方误差 (10^{-1})	4.63	21.89	4.15	2.45	1.0	5.04	0.95	0.53
	平均绝对百分比误差 (10^{-1})	2.51	1.61	1.69	3.7	1.24	0.81	0.82	1.8
	覆盖率（%）	95.5	94.6	95.0	94.4	94.9	95.0	95.0	93.9
MM2	偏差 (10^{-2})	6.62	19.42	7.38	4.6	1.7	2.64	1.05	1.64
	均方误差 (10^{-1})	3.53	13.71	2.72	2.17	0.81	2.9	0.61	0.49
	平均绝对百分比误差 (10^{-1})	2.26	1.3	1.36	3.6	1.12	0.61	0.66	1.73
	覆盖率（%）	94.1	93.6	93.9	94.2	94.4	94.9	95.8	93.9
MM3	偏差 (10^{-2})	6.62	19.42	7.38	4.6	1.7	2.64	1.05	1.64
	均方误差 (10^{-1})	3.53	13.71	2.72	2.17	0.81	2.9	0.61	0.49
	平均绝对百分比误差 (10^{-1})	2.26	1.3	1.36	3.6	1.12	0.61	0.66	1.73
	覆盖率（%）	95.2	94.3	95.3	94.7	94.3	95.0	96.1	93.9
MM4	偏差 (10^{-2})	3.56	9.23	3.11	3.03	1.0	0.26	0.04	1.29
	均方误差 (10^{-1})	3.27	12.52	2.55	2.05	0.8	2.85	0.61	0.48
	平均绝对百分比误差 (10^{-1})	2.2	1.25	1.33	3.53	1.12	0.6	0.66	1.72
	覆盖率（%）	95.6	95.4	94.9	94.0	94.1	96.0	96.3	94.4
BE1	偏差 (10^{-2})	-26.35	-111.25	-44.04	-10.55	-8.27	-34.25	-13.3	-3.12
	均方误差 (10^{-1})	2.02	17.48	3.21	1.03	0.57	3.35	0.64	0.32
	平均绝对百分比误差 (10^{-1})	1.87	1.66	1.61	2.58	0.95	0.68	0.68	1.42
	覆盖率（%）	89.2	74.6	81.3	91.1	92.8	88.7	90.2	92.8
BE2	偏差 (10^{-2})	-30.66	-115.58	-45.6	-14.83	-9.53	-35.49	-13.79	-4.36
	均方误差 (10^{-1})	2.24	18.4	3.32	1.12	0.59	3.43	0.65	0.33
	平均绝对百分比误差 (10^{-1})	2.0	1.71	1.65	2.73	0.97	0.69	0.69	1.44
	覆盖率（%）	89.2	74.6	81.3	91.1	92.8	88.7	90.2	92.8

表 C.2：当生成过程的真实值为 α，且生成过程的 $\alpha=(2,\ 7,\ 3,\ 1)$，样本数量为 $n=50$ 和 $n=200$ 时，六种方法的偏差估计、均方误差、平均绝对误差和覆盖率。估计值通过蒙特卡洛方法计算，进行了 1，000 次迭代，详见第 5.1 节。

		$n=50$				$n=1000$			
方法	评估	α_1	α_2	α_3	α_4	α_1	α_2	α_3	α_4
MM1	偏差 (10^{-2})	1.56	1.34	5.19	3.6	0.42	0.54	1.5	1.38
	均方误差 (10^{-1})	0.63	0.69	2.5	1.67	0.16	0.17	0.52	0.4
	平均绝对百分比误差 (10^{-1})	2.73	2.29	1.87	2.05	1.42	1.15	0.9	1.04
	覆盖率 (%)	93.96	94.31	94.66	95.35	94.6	94.5	96.4	94.4
MM2	偏差 (10^{-2})	2.12	1.77	5.57	4.29	0.43	0.55	1.39	1.33
	均方误差 (10^{-1})	0.55	0.55	1.48	1.21	0.14	0.14	0.32	0.29
	平均绝对百分比误差 (10^{-1})	2.66	2.09	1.5	1.81	1.34	1.05	0.7	0.9
	覆盖率 (%)	94.03	95.14	94.52	94.38	92.7	93.1	95.8	94.8
MM3	偏差 (10^{-2})	2.12	1.77	5.57	4.29	0.43	0.55	1.39	1.33
	均方误差 (10^{-1})	0.55	0.55	1.48	1.21	0.14	0.14	0.32	0.29
	平均绝对百分比误差 (10^{-1})	2.66	2.09	1.5	1.81	1.34	1.05	0.7	0.9
	覆盖率 (%)	93.75	94.73	95.07	93.2	93.0	93.9	94.9	95.0
MM4	偏差 (10^{-2})	1.11	0.48	2.57	2.05	0.16	0.2	0.59	0.73
	均方误差 (10^{-1})	0.53	0.53	1.37	1.14	0.14	0.14	0.31	0.29
	平均绝对百分比误差 (10^{-1})	2.63	2.07	1.47	1.77	1.33	1.06	0.7	0.89
	覆盖率 (%)	94.73	94.45	95.49	93.62	93.3	94.0	95.8	94.4
BE1	偏差 (10^{-2})	2.31	-2.46	-4.37	0.73	0.65	-0.66	-1.3	0.63
	均方误差 (10^{-1})	0.34	0.39	1.05	0.77	0.11	0.12	0.28	0.23
	平均绝对百分比误差 (10^{-1})	2.1	1.78	1.31	1.46	1.19	0.97	0.66	0.79
	覆盖率 (%)	95.84	95.07	94.86	95.14	94.7	94.1	94.8	95.2
BE2	偏差 (10^{-2})	1.04	-3.3	-6.07	-0.96	0.31	-0.83	-1.71	0.19
	均方误差 (10^{-1})	0.35	0.4	1.06	0.77	0.11	0.12	0.28	0.23
	平均绝对百分比误差 (10^{-1})	2.12	1.8	1.31	1.47	1.19	0.97	0.67	0.79
	覆盖率 (%)	95.84	95.07	94.86	95.14	94.7	94.1	94.8	95.2

表 C.3：当生成过程的真实值为α时，六种方法的偏差估计、均方误差、平均绝对误差和覆盖率 $\alpha = (0.7, 0.9, 2, 1.5)$ 下，$n = 50$ 和 $n = 200$。估计值通过蒙特卡洛方法进行计算，迭代 1,000 次，如第 5.1 节所述。

方法	评估	$\mathbb{E}[X]$	$\mathbb{E}[Y]$	方差(X)	方差(Y)	协方差$(X,\ Y)$
MM1	偏差 (10^{-2})	-0.163	-0.221	0.008	0.008	0.399
	均方误差 (10^{-2})	0.089	0.096	0.004	0.004	2.015
	平均绝对百分比误差 (10^{-1})	0.475	0.494	1.202	1.196	11.85
MM2	偏差 (10^{-2})	-0.163	-0.221	-0.049	-0.049	0.399
	均方误差 (10^{-2})	0.089	0.096	0.002	0.002	2.015
	平均绝对百分比误差 (10^{-1})	0.475	0.494	0.889	0.888	11.85
MM3	偏差 (10^{-2})	-0.163	-0.221	-0.049	-0.049	0.399
	均方误差 (10^{-2})	0.089	0.096	0.002	0.002	2.015
	平均绝对百分比误差 (10^{-1})	0.475	0.494	0.889	0.888	11.85
MM4	偏差 (10^{-2})	-0.163	-0.221	0.001	0.001	0.399
	均方误差 (10^{-2})	0.089	0.096	0.002	0.002	2.015
	平均绝对百分比误差 (10^{-1})	0.475	0.494	0.873	0.872	11.85
BE1	偏差 (10^{-2})	-0.153	-0.22	-0.075	-0.074	-2.817
	均方误差 (10^{-2})	0.084	0.09	0.002	0.002	1.214
	平均绝对百分比误差 (10^{-1})	0.458	0.481	0.732	0.732	9.076
BE2	偏差 (10^{-2})	-0.154	-0.223	-0.036	-0.036	-2.69
	均方误差 (10^{-2})	0.085	0.091	0.002	0.002	1.252
	平均绝对百分比误差 (10^{-1})	0.46	0.483	0.727	0.727	9.225

表 C.4：当 $\mu = (0,\ 0)$、$\sigma = [[1, 0.1], [0.1, 1]]$ 和 $n = 50$ 时，比较真实矩与双变量贝塔模型的估计值，给出偏差估计、均方误差和平均绝对百分比误差。

方法	评估	$\mathbb{E}[X]$	$\mathbb{E}[Y]$	方差(X)	方差(Y)	协方差(X, Y)
MM1	偏差 (10^{-2})	3.398	3.937	0.029	2.709	18.1
	均方误差 (10^{-2})	0.24	0.226	0.011	0.082	3.376
	平均绝对百分比误差 (10^{-1})	1.218	1.353	1.354	8.13	2.494
MM2	偏差 (10^{-2})	3.398	3.937	-1.743	0.994	18.1
	均方误差 (10^{-2})	0.24	0.226	0.035	0.015	3.376
	平均绝对百分比误差 (10^{-1})	1.218	1.353	2.826	3.099	2.494
MM3	偏差 (10^{-2})	-0.238	0.19	-1.765	0.932	26.5
	均方误差 (10^{-2})	0.139	0.071	0.036	0.013	7.091
	平均绝对百分比误差 (10^{-1})	0.903	0.694	2.863	2.948	3.652
MM4	偏差 (10^{-2})	8.151	8.596	-1.358	1.396	5.875
	均方误差 (10^{-2})	0.789	0.831	0.024	0.024	0.568
	平均绝对百分比误差 (10^{-1})	2.488	2.839	2.235	4.209	0.862
BE1	偏差 (10^{-2})	-2.674	2.059	-2.294	0.699	34.715
	均方误差 (10^{-2})	0.189	0.092	0.056	0.008	12.162
	平均绝对百分比误差 (10^{-1})	1.088	0.813	3.717	2.203	4.784
BE2	偏差 (10^{-2})	-2.843	1.894	-2.275	0.721	33.919
	均方误差 (10^{-2})	0.198	0.086	0.055	0.008	11.616
	平均绝对百分比误差 (10^{-1})	1.121	0.781	3.686	2.261	4.674

表 C.5：当 $\mu = (-1, -1)$ 且 $\sigma = [[2.25, -1.2], [-1.2, 1]]$ 和$n = 50$ 时，双变量贝塔模型的真实矩与估计矩之间的偏差估计、均方误差和平均绝对百分比误差。

D S 的渐近分布

根据Arellano-Valle 等人(2021) 中的命题 2,

$$\sqrt{n}\begin{pmatrix} \bar{X}_n - m_1 \\ S^2_{X,n} - v_1 \\ \bar{Y}_n - m_2 \\ S^2_{Y,n} - v_2 \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

对于协方差矩阵 Σ。注意，对于所有 $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ ， ∇g 均存在且连续。根据多元德尔塔方法,

$$\sqrt{n}G_n \xrightarrow{d} N\left(0, \nabla g(m_1, v_1, m_2, v_2)^T \Sigma \nabla g(m_1, v_1, m_2, v_2)\right),$$

因为 $g(m_1, v_1, m_2, v_2) = 0$。使用论文的记号，对于我们的应用， $d = 2, p = 2$, $D_{11} = X_n - m_1$, $D_{21} = Y_n - m_2$,$S_{12} = (n-1)S^2_{X,n}/n$, $S_{22} = (n-1)S^2_{Y,n}/n$, $\kappa_{11} = 0$, $\kappa_{21} = 0$, $\kappa_{12} = v_1$ 以及 $\kappa_{22} = v_2$。根据命题，由于 $\mathbb{E}[(X - m_1)^4]$ 和 $\mathbb{E}[(Y - m_2)^4]$ 均定义良好。

$$\sqrt{n}\begin{pmatrix} D_{11} \\ S_{12} \\ D_{21} \\ S_{22} \end{pmatrix} \xrightarrow{d} N_4\left(0, C\mathcal{K}C^T\right),$$

其中 $C = I_4$ 和 $\mathcal{K} =$

$$\begin{pmatrix} v_1 & \mathrm{Cov}(X, (X-m_1)^2) & \rho\sqrt{v_1 v_2} & \mathrm{Cov}(X, (Y-m_2)^2) \\ \mathrm{Cov}((X-m_1)^2, X) & \mathrm{Var}\,(X-m_1)^2 & \mathrm{Cov}((X-m_1)^2, Y) & \mathrm{Cov}((X-m_1)^2, (Y-m_2)^2) \\ \rho\sqrt{v_1 v_2} & \mathrm{Cov}(Y, (X-m_1)^2) & v_2 & \mathrm{Cov}(Y, (Y-m_2)^2) \\ \mathrm{Cov}((Y-m_2)^2, X) & \mathrm{Cov}((Y-m_2)^2, (X-m_1)^2) & \mathrm{Cov}((Y-m_2)^2, Y) & \mathrm{Var}\,(Y-m_2)^2. \end{pmatrix}$$

此外， $S^2_{X,n} \to S_{12}$ 和 $S^2_{Y,n} \to S_{22}$。

根据 ∇g 的连续性以及统计量 $\bar{X}_n$、 $\bar{Y}_n$、 $S^2_{X,n}$ 和 $S^2_{Y,n}$ 到 m_1、 m_2、 v_1 和v_2的相合性， $\nabla g(\bar{X}_n$、$S^2_{X,n}$、$\bar{Y}_n$、$S^2_{Y,n})$ 是 $\nabla g(m_1$、v_1、m_2、$v_2)$ 的相合估计量。因此，我们可以计算

$\hat{}nG_n$的极限分布方差的相合估计量。令 $\hat{}\sigma^2$ 为该估计量。根据斯卢茨基定理，我们得出

$$\sqrt{n}\frac{G_n}{\hat{\sigma}} \to N(0,1).$$

E 关于积分的说明

$(X,\ Y)$ 的密度为 $f_{X,\ Y}(x,\ y)$，如方程 (3) 所示；它在 $\mathbb{R}^2$ 中勒贝格零测度的集合上可能未定义。例如，在网格上作图时，这些集合可能很重要。本节将说明其中一个这样的集合。

若 $\alpha_i \geq 1,\ i = 1,\ \ldots\ldots,\ 4$，则显然对每个 $x,\ y \in [0,1]$，积分都有明确定义。令 $0 < \alpha_2 = \alpha_3 = a \leq 0.5$ 和 $x = y < 0.5$。于是

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{B(\boldsymbol{\alpha})}\int_0^x u_1^{\alpha_1-1}(x-u_1)^{a-1}(x-u_1)^{a-1}(1-2x+u_1)^{\alpha_4-1}\,du_1 \\ &= \frac{1}{B(\boldsymbol{\alpha})}\int_0^{x/2} u_1^{\alpha_1-1}(x-u_1)^{2a-2}(1-2x+u_1)^{\alpha_4-1}\,du_1 + \\ &\quad + \frac{1}{B(\boldsymbol{\alpha})}\int_{x/2}^x u_1^{\alpha_1-1}(x-u_1)^{2a-2}(1-2x+u_1)^{\alpha_4-1}\,du_1 \end{aligned}$$

请注意，第一个积分定义良好且非负。另一方面，第二个积分 al 未定义 ned:

$$\begin{aligned} &\int_{x/2}^x u_1^{\alpha_1-1}(x-u_1)^{2a-2}(1-2x+u_1)^{\alpha_4-1}\,du_1 \\ &\quad \geq \int_{x/2}^x \min\left(\left(\frac{x}{2}\right)^{\alpha_1-1}, x^{\alpha_1-1}\right)(x-u_1)^{2a-2} \\ &\qquad\qquad \times \min\left(\left(1-\frac{3}{2}x\right)^{\alpha_4-1}, (1-x)^{\alpha_4-1}\right)du_1 \\ &\quad = K(x)\int_0^{x/2} v^{2a-2}\,dv \\ &\quad = \begin{cases} \dfrac{K(x)}{2a-1}\lim_{t\to 0^+}\left[(x/2)^{2a-1} - t^{2a-1}\right] & \text{if } a < 0.5 \\ K(x)\lim_{t\to 0^+}\left[\log(x/2) - \log(t)\right] & \text{if } a = 0.5 \end{cases} \\ &\quad \to +\infty, \end{aligned}$$

其中，$K(x)$ 是 x 的函数。

根据这一发散现象，我们得出结论：如果 $0 < \alpha_2 = \alpha_3 \leq 0.5$ 且 $x = y < 0.5$，$f_{X,\ Y}(x,\ y)$ 未定义。请注意，如果 $x = y \geq 0.5$，仍会出现发散问题，因为当u_1 趋近于 x时就会出现这些问题。类似的计算表明，如果 $x + y = 1$ 且 $0 < \alpha_1 = \alpha_4 \leq 0.5$，密度也未定义。更一般地说，如果 $\alpha_1 + \alpha_4 \leq 1$ 且 $x + y = 1$，则 $f_{X,\ Y}(x,\ y)$ 未定义；$\alpha_2 + \alpha_3 \leq 1$ 且$x = y$。