分类和有序随机变量的多元概率分布

Takashi Arai*

山形大学理学院，日本山形 *990-8560*

arXiv:2304.00617v1 [stat.ME] 2 Apr 2023

摘要

我们提出一种适用于分类随机变量和有序随机变量的多元概率分布。为此，我们将格拉斯曼分布与分类变量和有序变量的虚拟变量编码结合起来。为了实现分类变量和有序变量所需的虚拟变量共现概率，我们提出了一种格拉斯曼分布的简约参数化方法，以确保概率分布的正性。作为所提出分布的一项应用，我们针对分类变量和有序变量开发了一种因子分析方法，并使用真实数据集验证了该模型的有效性。

* takashi-arai@sci.kj.yamagata-u.ac.jp

I. 引言

在数据分析中，经常会遇到分类变量和有序变量等定性数据。例如，性别和职业等人口统计属性、李克特量表中的问卷回答 [1], ，以及 DNA 和氨基酸序列中的突变是否存在。当定性变量作为因变量出现时，可以采用多类别逻辑回归或概率单位回归 [2], ；然而，难以为定性变量构建概率生成模型。定性变量生成模型的一个例子是使用分类分布的朴素贝叶斯分类器 [2], ，但该模型假设变量之间相互独立，无法考虑相关性。一种可能的方法是通过潜变量来表示相关性，例如二元因子分析 [3]和指数族主成分分析 [4, 5]。然而，通常引入潜变量的缺点是，所诱导的概率分布无法用解析形式表示。这一缺点使模型参数的估计变得困难，因此不得不求助于计算量大且耗时的近似方法。因此，定量处理定性数据是一项具有挑战性的任务。

处理多元分类与序数变量最简单的方法，是将它们视为连续变量，忽略其定性本质。例如，在五点量表的问卷回答中，响应变量 $\{1/5, 2/5, \ldots, 1\}$ 被视为连续数值变量。这种处理定性变量的方法称为量化方法。然而，这种量化方法将顺序尺度的变量视为比率尺度，也就是说，粗略地假设序数响应之间的间隔相等。因此，这种量化方法因缺乏审慎考量而受到批评。定性数据难以进行定量处理的原因之一，在于连续变量可以定义算术运算，而定性变量则无法进行加、减、乘、除等算术运算。对定性数据进行合理分析，往往局限于对列联表进行统计检验，例如费舍尔精确检验或 Cochran-Armitage检验。因此，要为定性数据建立合理的统计模型，就需要一种能够保留测量尺度本质的概率分布。

在分类变量中，多元二元随机变量可由格拉斯曼分布表示 [6]。此外，最近还开发出了一种用于建模相关性的概率分布

，用于描述二元变量与连续变量之间的关系 [7]。格拉斯曼分布具有与多元正态分布相似的良好理论性质；与传统的多元伯努利分布伊辛模型相比，它在计算配分函数时无需显式地对所有可能状态求和，因此具有计算优势。本文提出一种方法，将分类变量和序数变量进行虚拟变量编码，并结合格拉斯曼分布来表示多元分类与有序随机变量。

本文结构如下。第二节总结所提出的分布，并说明该分布为何能够再现分类变量和有序变量所需的虚拟变量共现概率。第三节提出一种简约参数化，以确保格拉斯曼分布的概率为正。该参数化是对角占优参数化的一种变体，也是所提出分布所必需的。第四节将因子分析作为所提出概率分布的一种应用，应用于分类和有序随机变量。通过分析一个真实数据集，展示所提出因子分析的有效性。文中还给出了双标图可视化及其解释。第五节为结论。

II. 所提出的分布

我们首先简要概述格拉斯曼分布。需要注意的是，本文中格拉斯曼分布的定义与先前研究中的格拉斯曼分布定义不同 [6, 7] ，区别在于虚拟变量的作用被反转了。此定义变更旨在使格拉斯曼分布的参数矩阵具有正则性。由于虚拟变量的作用类似于电流符号（正/负）的定义，因此这种变更是允许的，并非问题的本质。因此，我们概述本文所采用的格拉斯曼分布，记作 $\mathcal{G}^{(1)}$。

我们用列向量 $\mathbf{y}$ 表示 q 维虚拟变量，其中每个元素的取值为 0 或 1。也就是说，向量 $\mathbf{y}$ 是一个比特向量，其中每个元素的取值均为 0 或 1。我们将虚拟变量的所有索引集合记为 $R \equiv \{1, 2, \ldots, q\}$。虚拟变量的索引标签分为两部分，分别以下标 1 和 0 标记，用于表示

分别取值为 1 和 0 的变量。例如，二元变量 R 的索引标签分为子集 $R_1 \subseteq R$ 及其集合差 $R_0 = R \setminus R_1$。那么，格拉斯曼分布的联合分布表示如下:

$$\begin{aligned} p(\mathbf{y}_{R_1} = \mathbf{1}, \mathbf{y}_{R_0} = \mathbf{0}) =& \mathcal{G}^{(1)}(\mathbf{y}_{R_1} = \mathbf{1}, \mathbf{y}_{R_0} = \mathbf{0} \mid \Sigma), \\ =& \det \begin{bmatrix} I - \Sigma_{R_1 R_1} & \Sigma_{R_1 R_0} \\ -\Sigma_{R_0 R_1} & \Sigma_{R_0 R_0} \end{bmatrix} = \frac{\det(\Lambda_{R_1 R_1} - I)}{\det \Lambda}, \end{aligned} \tag{1}$$

其中，指标记法 $\Sigma_{R_1 R_1}$ 表示 Σ的子矩阵，而 I 是单位矩阵。

为表示边缘分布和条件分布，我们首先定义索引记法。我们将虚拟变量的完整索引集合划分为两个子集部分； $R =(S,\ T)$。这些索引集合中的元素个数分别用 q_S和 q_T表示，显然它们满足 $q_S + q_T = q$。然后，由索引子集 S 构成的子向量表示为 $\mathbf{y}_S$。虚拟变量的索引标签进一步划分为带下标 1 和 0 的两部分，分别对应取值为 1 和 0 的变量。例如，二元变量的索引标签 $S \subseteq R$ 划分为子集 $S_1 \subseteq S$ 及其集合差 $S_0 = S \setminus S_1$，其中子向量 $\mathbf{y}_{S_1}$ 和$\mathbf{y}_{S_0}$ 分别取值为 $\mathbf{y}_{S_1} = \mathbf{1}$ 和$\mathbf{y}_{S_0} = \mathbf{0}$。使用上述指标记法，边缘分布表示为

$$p(\mathbf{y}_T) = \mathcal{G}^{(1)}(\mathbf{y}_T \mid \Sigma_{TT}). \tag{2}$$

条件分布为

$$p(\mathbf{y}_S \mid \mathbf{y}_T) = \mathcal{G}^{(1)}(\mathbf{y}_S \mid \tilde{\Sigma}_{S|T}), \tag{3}$$

$$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{SS} & \tilde{\Sigma}_{ST} \\ \tilde{\Sigma}_{TS} & \tilde{\Sigma}_{TT} \end{bmatrix} \equiv \begin{bmatrix} \Sigma_{SS} & -\Sigma_{ST_1} & \Sigma_{ST_0} \\ \Sigma_{T_1 S} & I - \Sigma_{T_1 T_1} & \Sigma_{T_1 T_0} \\ \Sigma_{T_0 S} & -\Sigma_{T_0 T_1} & \Sigma_{T_0 T_0} \end{bmatrix}, \tag{4}$$

$$\begin{aligned} \tilde{\Sigma}_{S|T} \equiv& \tilde{\Sigma}_{SS} - \tilde{\Sigma}_{ST} \tilde{\Sigma}_{TT}^{-1} \tilde{\Sigma}_{TS}, \\ =& [\Lambda_{SS} - \Lambda_{ST_1} (\Lambda_{T_1 T_1} - I)^{-1} \Lambda_{T_1 S}]^{-1}, \end{aligned} \tag{5}$$

其中， Σ_{TT}^{-1}表示该子矩阵的逆矩阵， $\Sigma_{TT}^{-1} \equiv (\Sigma_{TT})^{-1}$， $\Sigma_{S|T}$是舒尔补。

虚拟变量的均值和协方差为

$$E[y_r] = 1 - \Sigma_{rr}, \tag{6}$$

$$\mathrm{Cov}[y_r\, y_s] = -\Sigma_{rs}\Sigma_{sr}, \quad (r, s \in R). \tag{7}$$

也就是说， Σ 的对角元素表示虚拟变量取值为 0 的边际概率。另一方面，当观察到所有其他条件变量均为 0 时，条件均值和协方差为

$$E[y_r \mid \mathbf{y}_{R\setminus r} = \mathbf{0}] = 1 - \frac{1}{\Lambda_{rr}} = \left(\frac{\Lambda_{rr}}{\Lambda_{rr} - 1}\right)^{-1}, \tag{8}$$

$$\mathrm{Cov}[y_r\, y_s \mid \mathbf{y}_{R\setminus\{r,s\}} = \mathbf{0}] = \frac{-\Lambda_{rs}\Lambda_{sr}}{(\Lambda_{rr}\Lambda_{ss} - \Lambda_{rs}\Lambda_{sr})^2}. \tag{9}$$

也就是说，当观察到所有其他条件变量均为 0 时， Λ 的对角元素表示虚拟变量取值为 0 的条件概率。

描述连续变量与虚拟变量之间相关性的概率分布汇总于附录 A。

A. 结果陈述

首先，我们用虚拟变量编码表示分类变量和有序变量。

对于分类变量，我们使用独热编码；对于具有 $q_C + 1$ 个类别 $\delta \in \{0, 1, 2, \ldots, q_C\}$ 的分类变量，我们用 q_C 个虚拟变量 $y_l \in \{0, 1\}$，(l =1，2、q_C) 表示，每个虚拟变量仅有一个取值为 1。也就是说，我们排除基准类别对应的虚拟变量。表 I 展示了一个包含四个类别的独热编码示例。

对于有序变量，我们采用类似于独热编码的虚拟变量编码。我们使用 q_O 个虚拟变量 $y_l \in \{0, 1\}$，($l = 1$，2，.....，q_O) 来表示具有 ($q_O + 1$) 个等级的有序变量 $\eta \in \{0, 1, 2, \ldots, q_O\}$；每个虚拟变量表示等于或大于一、二等的标记。例如，在四个等级的情况下，有序变量 $\eta = 2$ 可表示为 $y_1 = 1$、$y_2 = 1$ 和 $y_3 = 0$，因为它大于二且小于三。四个等级情况下的虚拟变量编码如表 II 所示。

请注意，对于分类变量和有序变量，单个虚拟变量本身没有意义，但多个虚拟变量的组合则有意义。二元变量可以视为具有两个类别的分类变量，也可以视为有两个

等级的有序变量。因此，我们将二元变量视为分类变量的一种特殊情况。

表 I. 类别虚拟变量。

δ	y_1	y_2	y_3
$\delta=0$	0	0	0
$\delta=1$	1	0	0
$\delta=2$	0	1	0
$\delta=3$	0	0	1

表 II. 有序虚拟变量。

η	y_1	y_2	y_3
$\eta=0$	0	0	0
$\eta=1$	1	0	0
$\eta=2$	1	1	0
$\eta=3$	1	1	1

考虑存在多个分类变量和有序变量的情况。首先，我们定义一种记法来统一表示这些变量。我们分别用 c 和 o表示分类变量和有序变量的数量。第 j个分类变量 δ_j 的数量记为 q_{C_j}。也就是说，分类变量 δ_j 取值于 $\delta_j \in\{0,1,2,\ldots,q_{C_j}\}$，($j=1$，2，……，$c$)。类似地，第 j个有序变量 η_j的数量记为 q_{O_j}。也就是说，有序变量 η_j取值于 $\eta_j \in \{0,1,2,\ldots,q_{O_j}\}$，($j$ =1，2，……，o)。分类变量 δ_j由虚拟变量的列向量$\mathbf{y}^{C_j}$编码，其中$\mathbf{y}^{C_j}$是一个q_{C_j}维比特向量，其中最多只有一个元素可以置为 1，其余元素均为 0。类似地，有序变量 η_j 由虚拟变量的列向量$\mathbf{y}^{O_j}$编码，其中

$$\mathbf{y}_{O_j}=[1,1,\ldots,1,0,0,\ldots,0]^T \tag{10}$$

是一个 q_{O_j} 维比特向量，其中所有取值为 1 的元素都排列在左侧。类别变量和有序变量的虚拟向量合并为一个虚拟向量 $\mathbf{y}$，

$$\mathbf{y}=[\mathbf{y}_C^T,\mathbf{y}_O^T]=[\mathbf{y}_{C_1}^T,\mathbf{y}_{C_2}^T,\ldots,\mathbf{y}_{C_c}^T,\mathbf{y}_{O_1}^T,\mathbf{y}_{O_2}^T,\ldots,\mathbf{y}_{O_o}^T]^T, \tag{11}$$

其中，组合虚拟向量 $\mathbf{y}$ 的维度数为 $q=q_C+q_O\equiv q_{C_1}+q_{C_2}+\cdots+q_{C_c}+q_{O_1}+q_{O_2}+\cdots+q_{O_o}$。我们用 R表示全体索引的集合 $\{1,2,\ldots,q\}$。对于索引集 R，我们分别用 C 和 O表示对应类别型变量和有序变量的索引集合。这些索引集满足 $O=R\setminus C$。这些索引集合被划分为各个变量的索引集合，即 $C=$(C_1，C_2，……，C_c) 和$O=$(O_1，O_2，……，O_o)。

上述类别型和有序型虚拟变量服从 Grassmann 分布，

$$p(\mathbf{y}) = \mathcal{G}^{(1)}(\mathbf{y} \mid \Sigma = \Lambda^{-1}), \tag{12}$$

其中，格拉斯曼分布的参数用以下简洁表达式表示：

$$\Lambda - I = \Psi^{-1} - I + W\Omega V^T, \tag{13}$$

其中， T 表示矩阵转置。参数 V 是一个 $q \times a$ 矩阵，其中正整数 $a \in \{1, 2, \ldots, q\}$ 表示辅助潜在维度的数量，并控制模型的可表示性；取值越大，模型的可表示性越强。Ω是一个 q 阶对角矩阵，其对角元素的取值介于零和一之间。 $q \times q$ 矩阵 Ψ 必须是一个分块对角矩阵，使得 $\Psi^{-1} - I$ 的每个块元素都是一个准对角矩阵，且具有非负主子式；其定义如下，这一要求确保类别型和有序型虚拟变量的某些共现概率恰好为零：

$$\Psi^{-1} - I = \begin{bmatrix} (\Psi^{-1} - I)_{CC} & O \\ O & (\Psi^{-1} - I)_{OO} \end{bmatrix}, \tag{14}$$

$$(\Psi^{-1} - I)_{CC} = \begin{bmatrix} (\Psi^{-1} - I)_{C_1C_1} & O & \cdots & O \\ O & (\Psi^{-1} - I)_{C_2C_2} & & \vdots \\ \vdots & & \ddots & O \\ O & \cdots & O & (\Psi^{-1} - I)_{C_cC_c} \end{bmatrix}, \tag{15}$$

$$(\Psi^{-1} - I)_{C_jC_j} = \begin{bmatrix} (\boldsymbol{\psi}^{(C_j)})^T \\ (\boldsymbol{\psi}^{(C_j)})^T \\ \vdots \\ (\boldsymbol{\psi}^{(C_j)})^T \end{bmatrix} = \begin{bmatrix} e^{b_1^{(C_j)}} & e^{b_2^{(C_j)}} & \cdots & e^{b_{qC_j}^{(C_j)}} \\ e^{b_1^{(C_j)}} & e^{b_2^{(C_j)}} & \cdots & e^{b_{qC_j}^{(C_j)}} \\ \vdots & \vdots & \cdots & \vdots \\ e^{b_1^{(C_j)}} & e^{b_2^{(C_j)}} & \cdots & e^{b_{qC_j}^{(C_j)}} \end{bmatrix}, \quad (j = 1, 2, \ldots, c), \tag{16}$$

$$\tag{17}$$

其中， $\boldsymbol{\psi}^{(C_j)}$ 和$\mathbf{b}^{(C_j)}$ 是q_{C_j}维列向量，并且

$$(\Psi^{-1}-I)_{OO}=\begin{bmatrix}(\Psi^{-1}-I)_{O_1O_1} & O & \cdots & O\\ O & (\Psi^{-1}-I)_{O_2O_2} & & \vdots\\ \vdots & & \ddots & O\\ O & \cdots & O & (\Psi^{-1}-I)_{O_oO_o}\end{bmatrix}, \tag{18}$$

$$\begin{aligned}(\Psi^{-1}-I)_{O_jO_j} =&L_{O_j}+\left[\boldsymbol{\psi}^{(O_j)},\ \mathbf{0},\ \ldots,\ \mathbf{0}\right]^T,\\ =&L_{O_j}+\begin{bmatrix}e^{b_1^{(O_j)}} & e^{b_1^{(O_j)}+b_2^{(O_j)}} & \cdots & \prod_{l=1}^{q_{O_j}}e^{b_l^{(O_j)}}\\ 0 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & 0\end{bmatrix},\quad (j=1,2,\ldots,o),\end{aligned} \tag{19}$$

其中 $\boldsymbol{\psi}^{(O_j)}$ 和 $\mathbf{b}^{(O_j)}$ 是 q_{O_j} 维列向量，并且

$$L_{O_j}=\begin{bmatrix}0 & 0 & 0 & \cdots & 0\\ -1 & 0 & 0 & \cdots & 0\\ 0 & -1 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0 & 0\\ 0 & 0 & 0 & -1 & 0\end{bmatrix} \tag{20}$$

是一个 q_{O_j} 阶下三角矩阵，其元素均为 **-1**。由于类别型和有序型虚拟变量的共现概率必须为零， $q\times a$ 矩阵 W必须具有以下形式：

$$W=\left[W_{CA}^T,\ W_{OA}^T\right]^T, \tag{21}$$

$$W_{CA}=\left[W_{C_1A}^T,\ W_{C_2A}^T,\ \ldots,\ W_{C_cA}^T\right]^T,\ W_{C_jA}=\left[\mathbf{w}^{(C_j)},\ \mathbf{w}^{(C_j)},\ \ldots,\ \mathbf{w}^{(C_j)}\right]^T,\ (j=1,2,\ldots,c), \tag{22}$$

$$W_{OA}=\left[W_{O_1A}^T,\ W_{O_2A}^T,\ \ldots,\ W_{O_oA}^T\right]^T,\quad W_{O_jA}=\left[\mathbf{w}^{(O_j)},\ \mathbf{0},\ \ldots,\ \mathbf{0}\right]^T,\quad (j=1,2,\ldots,o), \tag{23}$$

其中 W_{C_jA} 是一个 $q_{C_j}\times a$ 矩阵，$\mathbf{w}^{(C_j)}$ 是一个 a 维列向量。此外， W_{O_jA} 是一个 $q_{O_j}\times a$ 矩阵，$\mathbf{w}^{(O_j)}$ 是一个 a 维列向量。

格拉斯曼分布由显式模型参数 θ_m =(Ψ， W， V， Ω)以及辅助参数 θ_a =(C) =(C_{RR}， C_{RA}， C_{AR}， C_{AA})进行参数化。此

辅助参数不会直接影响模型的似然。$(q+a) \times (q+a)$ 参数矩阵 C 按如下方式进行分块：

$$C = \begin{bmatrix} C_{RR} & C_{RA} \\ C_{AR} & C_{AA} \end{bmatrix}, \tag{24}$$

并且必须是严格对角占优矩阵。此外，由模型参数组合表示的以下矩阵 B 必须是对角占优矩阵：

$$B = \begin{bmatrix} B_{RR} & B_{RA} \\ B_{AR} & B_{AA} \end{bmatrix} = \begin{bmatrix} (\Psi^{-1} - I + WV^T)C_{RR} - WC_{AR}, & (\Psi^{-1} - I + WV^T)C_{RA} - WC_{AA} \\ C_{AR} - V^T C_{RR}, & C_{AA} - V^T C_{RA} \end{bmatrix}. \tag{25}$$

这些对角占优和严格对角占优条件是格拉斯曼分布 $\Lambda - I$ 的参数矩阵成为 P_0-矩阵的必要条件。也就是说，格拉斯曼分布通过 $\theta = (\theta_m, \ \theta_a)$ 进行参数化。

B. 虚拟变量的共现概率

直接使用格拉斯曼分布对类别型和有序型虚拟变量进行建模并不合适。这是因为采用这种简单方法时，多个虚拟变量同时取值为 1 的概率，即共现概率，可能不为零。本节将说明，上一节中格拉斯曼分布的参数化如何使不希望出现的共现概率恰好为零。本文有时将虚拟变量取值为 1 的概率称为激发概率。

1. 分类变量的激发概率

在分类变量的独热编码中，多个虚拟变量不会同时开启，最多只有一个虚拟变量开启。由于虚拟变量的观测˜概率由参数矩阵 Σ的行列式表示，如公式（4）所示，因此，当观测到多个虚拟变量为 1 时，应对矩阵 Σ 进行参数化，使参数矩阵的某些行向量重合，以便

使共现概率为零。在四个类别的具体示例中，以下参数化方式可以实现所需的激发概率：

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{22}-1 & \Sigma_{33}-1 \\ \Sigma_{11}-1 & \Sigma_{22} & \Sigma_{33}-1 \\ \Sigma_{11}-1 & \Sigma_{22}-1 & \Sigma_{33} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{22}-1 & \Lambda_{33}-1 \\ \Lambda_{11}-1 & \Lambda_{22} & \Lambda_{33}-1 \\ \Lambda_{11}-1 & \Lambda_{22}-1 & \Lambda_{33} \end{bmatrix}. \tag{26}$$

在上述参数化中，只有表 I 所示的四种不同激发概率会变为非零。这四种概率可由以下分类分布表示：

$$p(\delta) = \mathrm{Cat}(\mathbf{y}|\mathbf{b}) \equiv \frac{\exp(\mathbf{y}^T\mathbf{b})}{1+\sum_{l=1}^{q_C}\exp(b_l)}, \quad b_l = \log(\Lambda_{ll}-1). \tag{27}$$

2. 有序变量的激发概率

在有序变量的虚拟变量编码中，处于激活状态的虚拟向量元素会像公式（10）中那样靠左排列。这意味着，虚拟变量的较低等级为零而较高等级为一的共现概率必须恰好为零，例如，$y_{l-1}=0$ 和 $y_l=1$ 不能同时出现。我们来考虑一个具体的四个等级示例，即 $q_O=3$。满足上述共现概率约束的参数化可表示如下：

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{22}-1 & \Sigma_{33}-1 \\ \Sigma_{11} & \Sigma_{22} & \Sigma_{33}-1 \\ \Sigma_{11} & \Sigma_{22} & \Sigma_{33} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{28}$$

在上述参数化中，只有表 II 所示的四种不同激发概率会变为非零。这四种概率由如下定义的序数分布表示：

$$p(\eta) = \mathrm{Ord}(\mathbf{y}|\mathbf{b}) \equiv \frac{\exp(\mathbf{y}^T\mathbf{b})}{1+\sum_{l=1}^{q_O}\prod_{m=1}^{l}\exp(b_m)}, \quad \Lambda_{11}-1=e^{b_1},\ \Lambda_{12}=e^{b_1+b_2},\ \Lambda_{13}=e^{b_1+b_2+b_3}. \tag{29}$$

我们将形如式（26、28）所示的矩阵 $\Lambda - I$称为具有非负主子式的准对角矩阵。

3. 分类变量和序数变量的混合

当分类变量与等级数为四的序数变量混合时，上一节的结果可以用如下分块矩阵概括:

$$\Sigma=\begin{bmatrix}\Sigma_{CC} & \Sigma_{CO}\\ \Sigma_{OC} & \Sigma_{OO}\end{bmatrix}=\begin{bmatrix}\Sigma_{11} & \Sigma_{22}-1 & \Sigma_{33}-1 & \Sigma_{1O}\\ \Sigma_{11}-1 & \Sigma_{22} & \Sigma_{33}-1 & \Sigma_{1O}\\ \Sigma_{11}-1 & \Sigma_{22}-1 & \Sigma_{33} & \Sigma_{1O}\\ \Sigma_{O1} & \Sigma_{O2} & \Sigma_{O3} & \Sigma_{OO}\end{bmatrix}=\begin{bmatrix}\Sigma_{CC} & \Sigma_{C4} & \Sigma_{C5} & \Sigma_{C6}\\ \Sigma_{4C} & \Sigma_{44} & \Sigma_{55}-1 & \Sigma_{66}-1\\ \Sigma_{4C} & \Sigma_{44} & \Sigma_{55} & \Sigma_{66}-1\\ \Sigma_{4C} & \Sigma_{44} & \Sigma_{55} & \Sigma_{66}\end{bmatrix}, \tag{30}$$

$$\Lambda-I=\begin{bmatrix}\Lambda_{CC}-I & \Lambda_{CO}\\ \Lambda_{OC} & \Lambda_{OO}-I\end{bmatrix}=\begin{bmatrix}\Lambda_{11}-1 & \Lambda_{22}-1 & \Lambda_{33}-1 & \Lambda_{1O}\\ \Lambda_{11}-1 & \Lambda_{22}-1 & \Lambda_{33}-1 & \Lambda_{1O}\\ \Lambda_{11}-1 & \Lambda_{22}-1 & \Lambda_{33}-1 & \Lambda_{1O}\\ \Lambda_{O1} & \Lambda_{O2} & \Lambda_{O3} & \Lambda_{OO}\end{bmatrix}=\begin{bmatrix}\Lambda_{CC} & \Lambda_{C4} & \Lambda_{C5} & \Lambda_{C6}\\ \Lambda_{4C} & \Lambda_{44} & \Lambda_{45} & \Lambda_{46}\\ \mathbf{0}^T & -1 & 0 & 0\\ \mathbf{0}^T & 0 & -1 & 0\end{bmatrix}. \tag{31}$$

模型参数 $\Lambda-I$ 应表示为

$$\Lambda-I=\begin{bmatrix}(\Lambda-I)_{CR}\\ (\Lambda-I)_{OR}\end{bmatrix},\ (\Lambda-I)_{CR}=\begin{bmatrix}(\boldsymbol{\lambda}^{(C)})^T\\ (\boldsymbol{\lambda}^{(C)})^T\\ (\boldsymbol{\lambda}^{(C)})^T\end{bmatrix},\ (\Lambda-I)_{OR}=\begin{bmatrix}(\boldsymbol{\lambda}^{(O)})^T\\ \mathbf{0}^T\\ \mathbf{0}^T\end{bmatrix}+\begin{bmatrix}O, & L\end{bmatrix}, \tag{32}$$

其中， O 是所有元素均为零的矩阵， L 是以下下三角矩阵,

$$L=\begin{bmatrix}0 & 0 & 0\\ -1 & 0 & 0\\ 0 & -1 & 0\end{bmatrix}, \tag{33}$$

并且 $\boldsymbol{\lambda}^{(C)}$ 和 $\boldsymbol{\lambda}^{(O)}$ 是六维列向量。

多个类别变量和有序变量的情形也可以用分块矩阵表示。事实上，我们可以看到, $\Lambda-I$ 可以按如下方式进行参数化:

$$(\Lambda-I)_{CR}=\begin{bmatrix}(\Lambda-I)_{C_1R}\\ (\Lambda-I)_{C_2R}\\ \vdots\\ (\Lambda-I)_{C_cR}\end{bmatrix},\ (\Lambda-I)_{C_jR}=\begin{bmatrix}(\boldsymbol{\lambda}^{(C_j)})^T\\ (\boldsymbol{\lambda}^{(C_j)})^T\\ \vdots\\ (\boldsymbol{\lambda}^{(C_j)})^T\end{bmatrix},\quad (j=1,2,\ldots,c), \tag{34}$$

$$(\Lambda - I)_{OR} = \begin{bmatrix} (\Lambda - I)_{O_1R} \\ (\Lambda - I)_{O_2R} \\ \vdots \\ (\Lambda - I)_{O_oR} \end{bmatrix} = \begin{bmatrix} (\Lambda - I)_{O_1C},\ (\Lambda - I)_{O_1O} \\ (\Lambda - I)_{O_2C},\ (\Lambda - I)_{O_2O} \\ \vdots, \qquad \vdots \\ (\Lambda - I)_{O_oC},\ (\Lambda - I)_{O_oO} \end{bmatrix} = \begin{bmatrix} \tilde{\Lambda}_{O_1R} \\ \tilde{\Lambda}_{O_2R} \\ \vdots \\ \tilde{\Lambda}_{O_oR} \end{bmatrix} + \begin{bmatrix} O & L_{O_1} & O & \cdots & O \\ O & O & L_{O_2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & O \\ O & O & \cdots & O & L_{O_o} \end{bmatrix}, \tag{35}$$

$$\tilde{\Lambda}_{O_jR} = \left[\boldsymbol{\lambda}^{(O_j)},\ \mathbf{0},\ \ldots,\ \mathbf{0} \right]^T,\ \ (j = 1, 2, \ldots, o), \tag{36}$$

其中， L_{O_j} 是一个 q_{O_j} 阶下三角矩阵，其元素为式 (20) 中定义的负零。

III. 格拉斯曼分布的简约参数化

上一节中格拉斯曼分布的参数化无法通过传统的对角占优参数化 [6] 实现。因此，本节提出一种适用于格拉斯曼分布的新型简约参数化。

A. 简约参数化的推导

简约参数化是一种受连续变量传统因子分析思想启发的参数化方法。我们引入辅助虚拟变量的潜在维度，并用分块矩阵表示模型参数矩阵，该矩阵应为一个 P_0-矩阵。我们将观测虚拟变量和潜在虚拟变量的完整索引集合分别记为 R 和 A，并将这些维度分别记为 q 和 a。观测变量和潜变量合并后的完整索引集合记为 $K = (R,\ A)$。于是，具有完整索引的模型参数 $\Lambda - I$ 矩阵定义为

由以下分块上三角矩阵的乘积表示:

$$
\begin{aligned}
(\Lambda - I)_{KK} &= \begin{bmatrix} I & O \\ O & (\Omega^{-1} - I)^{1/2} \end{bmatrix} \begin{bmatrix} \Psi^{-1} - I & -W \\ O & I \end{bmatrix} \begin{bmatrix} I & O \\ -V^T & I \end{bmatrix} \begin{bmatrix} I & O \\ O & (\Omega^{-1} - I)^{1/2} \end{bmatrix}, \\
&= \begin{bmatrix} I & O \\ O & (\Omega^{-1} - I)^{1/2} \end{bmatrix} \begin{bmatrix} \Psi^{-1} - I + WV^T & -W \\ -V^T & I \end{bmatrix} \begin{bmatrix} I & O \\ O & (\Omega^{-1} - I)^{1/2} \end{bmatrix}, \\
&= \begin{bmatrix} \Psi^{-1} - I + WV^T, & -W(\Omega^{-1} - I)^{1/2} \\ -(\Omega^{-1} - I)^{1/2} V^T & \Omega^{-1} - I \end{bmatrix}, \qquad (37)
\end{aligned}
$$

$$
\Sigma_{KK} = \begin{bmatrix} \Lambda_{R|A}^{-1}, & \Lambda_{R|A}^{-1} W \Omega^{1/2} (I - \Omega)^{1/2} \\ \Omega^{1/2} (I - \Omega)^{1/2} V^T \Lambda_{R|A}^{-1}, & \Omega + \Omega^{1/2} (I - \Omega)^{1/2} V^T \Lambda_{R|A}^{-1} W \Omega^{1/2} (I - \Omega)^{1/2} \end{bmatrix}, \qquad (38)
$$

$$
\Lambda_{R|A} = \Psi^{-1} + W \Omega V^T, \qquad (39)
$$

其中，Ω 是一个对角矩阵，其对角元素满足 $0 \leq \text{diag}(\Omega) \leq 1$。 $\Psi^{-1} - I$是由准对角矩阵组成的分块对角矩阵，其主子式非负，如第 II B 节所定义；它是一个实现类别型和有序型虚拟变量共现概率的方阵。在上述参数化中，观测变量之间以及潜变量之间均满足条件独立性。事实上，当观测变量全部取值为 1 时，给定观测变量的潜变量格拉斯曼分布的模型参数为 $p(\mathbf{y}_A \mid \mathbf{y}_R = \mathbf{1}) = \mathcal{G}^{(1)}(\mathbf{y}_A \mid \Omega)$，这意味着给定观测变量时潜变量之间相互独立。另一方面，当潜变量全部取值为 0 时，给定潜变量的观测变量格拉斯曼分布的模型参数为 $p(\mathbf{y}_R \mid \mathbf{y}_A = \mathbf{0}) = \mathcal{G}^{(1)}(\mathbf{y}_R \mid \Psi)$，这再次意味着给定潜变量时观测变量之间相互独立。

P_0 矩阵可表示为形式为 BC^{-1} [8] 的行对角占优矩阵 B 和 C 的乘积。由于乘以非负对角矩阵不会改变 P_0 矩阵的性质，因此我们假设式 (37) 右侧的中间部分分解为行对角占优矩阵 B 与严格行对角占优

矩阵 C 的乘积,

$$\begin{aligned}
&\begin{bmatrix} \Psi^{-1} - I + WV^T & -W \\ -V^T & I \end{bmatrix}, \\
&= \begin{bmatrix} B_{RR} & B_{RA} \\ B_{AR} & B_{AA} \end{bmatrix} \begin{bmatrix} C_{RR} & C_{RA} \\ C_{AR} & C_{AA} \end{bmatrix}^{-1}, \\
&= \begin{bmatrix} B_{RR} & B_{RA} \\ B_{AR} & B_{AA} \end{bmatrix} \begin{bmatrix} C_{RR}^{-1} + C_{RR}^{-1} C_{RA} C_{A|R}^{-1} C_{AR} C_{RR}^{-1}, & -C_{RR}^{-1} C_{RA} C_{A|R}^{-1} \\ -C_{A|R}^{-1} C_{AR} C_{RR}^{-1}, & C_{A|R}^{-1} \end{bmatrix}, \\
&= \begin{bmatrix} B_{RR} C_{RR}^{-1} - (B_{RA} - B_{RR} C_{RR}^{-1} C_{RA}) C_{A|R}^{-1} C_{AR} C_{RR}^{-1}, & (B_{RA} - B_{RR} C_{RR}^{-1} C_{RA}) C_{A|R}^{-1} \\ B_{AR} C_{RR}^{-1} - (B_{AA} - B_{AR} C_{RR}^{-1} C_{RA}) C_{A|R}^{-1} C_{AR} C_{RR}^{-1}, & (B_{AA} - B_{AR} C_{RR}^{-1} C_{RA}) C_{A|R}^{-1} \end{bmatrix}. \qquad (40)
\end{aligned}$$

比较上式的第 A列，我们看到以下两个等式必须成立,

$$\begin{cases} (B_{RA} - B_{RR} C_{RR}^{-1} C_{RA}) = -W C_{A|R}, \\ (B_{AA} - B_{AR} C_{RR}^{-1} C_{RA}) = C_{A|R}. \end{cases} \qquad (41)$$

将上述等式代入式（40）的第 R列，我们得到以下两个关系,

$$\begin{cases} B_{RR} C_{RR}^{-1} + W C_{AR} C_{RR}^{-1} = \Psi^{-1} - I + WV^T, \\ B_{AR} C_{RR}^{-1} - C_{AR} C_{RR}^{-1} = -V^T, \end{cases} \Rightarrow \begin{cases} B_{RR} C_{RR}^{-1} = \Psi^{-1} - I + WV^T - W C_{AR} C_{RR}^{-1}, \\ B_{AR} C_{RR}^{-1} = C_{AR} C_{RR}^{-1} - V^T. \end{cases} \qquad (42)$$

因此，将$B_{RR} C_{RR}^{-1}$ 和$B_{AR} C_{RR}^{-1}$ 的上述表达式代入式（41），得到

$$\begin{cases} B_{RA} = (\Psi^{-1} - I + WV^T) C_{RA} - W C_{AA}, \\ B_{AA} = C_{AA} - V^T C_{RA}. \end{cases} \qquad (43)$$

根据上述微　　积分，对角占优矩阵 B 必须具有以下形式,

$$B = \begin{bmatrix} (\Psi^{-1} - I + WV^T) C_{RR} - W C_{AR}, & (\Psi^{-1} - I + WV^T) C_{RA} - W C_{AA} \\ C_{AR} - V^T C_{RR}, & C_{AA} - V^T C_{RA} \end{bmatrix}. \qquad (44)$$

对潜变量进行边缘化后，我们得到观测变量的格拉斯曼分布的简约模型参数化，$\theta = (\Psi,\ W,\ V,\ \Omega\ C)$:

$$\Sigma_{RR} = [\Psi^{-1} + W \Omega V^T]^{-1}, \qquad (45)$$

其中 B 满足行对角占优条件（式（44））， 而 C 满足严格行对角占优条件。

B. 类别变量与序数变量参数化的实现

在格拉斯曼分布的简约参数化中，模型参数 $\Lambda - I$ 表示为

$$\Lambda - I = \Psi^{-1} - I + W\Omega V^T. \tag{46}$$

如第 II B 节所述，类别变量和序数变量所需的参数化可通过对 Ψ 和 W施加约束来实现。

1. 仅包含类别变量

采用简约参数化，第 IIB1 节中的类别虚拟变量参数化

$$\Lambda - I = \begin{bmatrix} \Lambda_{11} - 1 & \Lambda_{22} - 1 & \Lambda_{33} - 1 \\ \Lambda_{11} - 1 & \Lambda_{22} - 1 & \Lambda_{33} - 1 \\ \Lambda_{11} - 1 & \Lambda_{22} - 1 & \Lambda_{33} - 1 \end{bmatrix} = \Psi^{-1} - I + W\Omega V^T, \tag{47}$$

可通过如下方式对 Ψ 和 W 施加约束来实现：

$$\Psi^{-1} - I = \begin{bmatrix} e^{b_1} & e^{b_2} & e^{b_3} \\ e^{b_1} & e^{b_2} & e^{b_3} \\ e^{b_1} & e^{b_2} & e^{b_3} \end{bmatrix}, \quad W = \begin{bmatrix} \mathbf{w}^T \\ \mathbf{w}^T \\ \mathbf{w}^T \end{bmatrix}, \tag{48}$$

其中，$\Psi^{-1} - I$ 是一个具有非负主子式的准对角矩阵，而$\mathbf{b}$ 和$\mathbf{w}$是三维列向量。

2. 仅序数变量

同样，第 II B 节 2 中有序虚拟变量的参数化

$$\Lambda - I = \begin{bmatrix} \Lambda_{11} - 1 & \Lambda_{12} & \Lambda_{13} \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} = \Psi^{-1} - I + W\Omega V^T, \tag{49}$$

可以通过对 Ψ 施加约束来实现，具体如下：

$$\Psi^{-1} - I = \begin{bmatrix} e^{b_1} & e^{b_1+b_2} & e^{b_1+b_2+b_3} \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} \mathbf{w}^T \\ \mathbf{0}^T \\ \mathbf{0}^T \end{bmatrix}, \tag{50}$$

其中，$\Psi^{-1} - I$ 是一个具有非负主子式的准对角矩阵，$\mathbf{b}$ 和$\mathbf{w}$是三维列向量。

3. 类别变量与有序变量的混合

类别变量与有序变量混合的参数化，可以通过在分块矩阵中表示 Ψ 和 W 来实现，如下所示：

$$\Lambda - I = \Psi^{-1} - I + W\Omega V^T = \begin{bmatrix} (\Psi^{-1} - I)_{CC} & O \\ O & (\Psi^{-1} - I)_{OO} \end{bmatrix} + \begin{bmatrix} W_{CA} \\ W_{OA} \end{bmatrix} \Omega V^T. \tag{51}$$

当单个类别变量和有序变量的等级数均为四时，可通过对 Ψ 和 W 施加如下约束来实现参数化：

$$(\Psi^{-1} - I)_{CC} = \begin{bmatrix} (\boldsymbol{\psi}^{(C)})^T \\ (\boldsymbol{\psi}^{(C)})^T \\ (\boldsymbol{\psi}^{(C)})^T \end{bmatrix} = \begin{bmatrix} e^{b_1^{(C)}} & e^{b_2^{(C)}} & e^{b_3^{(C)}} \\ e^{b_1^{(C)}} & e^{b_2^{(C)}} & e^{b_3^{(C)}} \\ e^{b_1^{(C)}} & e^{b_2^{(C)}} & e^{b_3^{(C)}} \end{bmatrix}, \tag{52}$$

$$(\Psi^{-1} - I)_{OO} = L + \begin{bmatrix} (\boldsymbol{\psi}^{(O)})^T \\ \mathbf{0}^T \\ \mathbf{0}^T \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} + \begin{bmatrix} e^{b_1^{(O)}} & e^{b_1^{(O)}+b_2^{(O)}} & e^{b_1^{(O)}+b_2^{(O)}+b_3^{(O)}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{53}$$

$$W_{CA} = \begin{bmatrix} \mathbf{w}^{(C)}, & \mathbf{w}^{(C)}, & \mathbf{w}^{(C)} \end{bmatrix}^T, \qquad W_{OA} = \begin{bmatrix} \mathbf{w}^{(O)}, & \mathbf{0}, & \mathbf{0} \end{bmatrix}^T. \tag{54}$$

多个类别变量和有序变量的情形，可以按照第 II A 节所述进行参数化。

C. 使用真实数据进行验证

在本节中，我们通过数值分析确认，所提出的简约参数化能够利用真实数据集，成功再现分类变量和序数虚拟变量的相关性及共现概率。本分析所使用的数据是读者数据

，来自 R（编程语言）的 catdata 包 [9]。读者数据包含女性针对某本特定女性杂志的阅读行为信息，由四个变量组成：该女性是否为定期读者？（是 $=1$，否 $=0$）、该女性是否工作？（是 $=1$，否 $=0$）、她们的年龄组（18–29 岁 $=1$，30–39 岁 $=2$，40–49 岁 $=3$），以及从 L1 到 L4 的教育程度。该数据集的样本量为 $N=941$。我们将年龄组视为分类变量，将教育程度视为序数变量。我们抽取了一个二元变量“工作状态”、一个分类变量“年龄类别”和一个序数变量“教育程度”进行分析。

模型参数通过最大似然估计得到。由于似然函数的值在辅助潜在维度数量为二时几乎达到饱和，因此将辅助潜在维度的数量选为二。估计得到的参数如下

$$\Lambda - I = \begin{bmatrix} 0.62 & 0.12 & -0.14 & -2.12 & -0.63 & -0.14 \\ -1.76 & 1.73 & 2.57 & -3.29 & -4.49 & -1.84 \\ -1.76 & 1.73 & 2.57 & -3.29 & -4.49 & -1.84 \\ 0.83 & -0.36 & -0.98 & 2.36 & 2.05 & 0.78 \\ 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 \end{bmatrix}, \tag{55}$$

$$\Sigma = \begin{bmatrix} 0.52 & -0.10 & 0.20 & 0.41 & -0.08 & -0.05 \\ 0.08 & 0.62 & -0.30 & 0.54 & 0.59 & 0.19 \\ 0.08 & -0.38 & 0.70 & 0.54 & 0.59 & 0.19 \\ -0.05 & -0.01 & 0.07 & 0.22 & -0.32 & -0.08 \\ -0.05 & -0.01 & 0.07 & 0.22 & 0.68 & -0.08 \\ -0.05 & -0.01 & 0.07 & 0.22 & 0.68 & 0.92 \end{bmatrix}, \tag{56}$$

其中数值四舍五入至小数点后两位以供展示。在此模型参数下，尽管对于虚拟变量而言，格拉斯曼分布有 $2^6 = 64$ 个激发概率，但所有不理想的共现概率都恰好为零，而理想的 12 个激发概率则得以保留。所提出的分布精确地再现了数据的经验均值。图1展示了经验相关性以及模型再现的相关性。可以看出，所提出的分布成功再现了类别型和有序型虚拟变量之间的相关性。

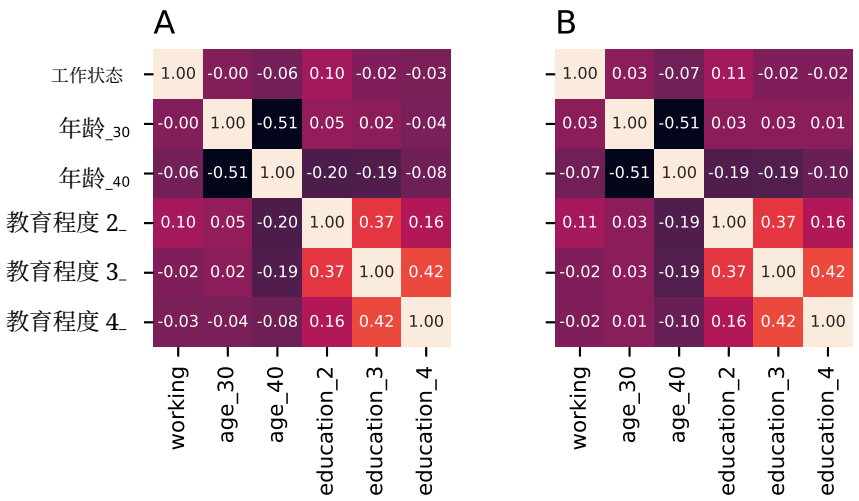


图1. 读者数据的皮尔逊相关矩阵，根据数据计算得出（A），并由模型复现（B）。

IV. 分类因子分析

作为多元分类与序数变量的一种应用，本节探讨针对类别型和有序变量的因子分析。通过为格拉斯曼分布设置参数 Λ，使其取值为分块对角矩阵，且 $\Lambda - I$ 的每个块元素都是具有非负主子式的准对角矩阵（定义见第 II B 节），即可实现所提出的因子分析。这种设置意味着，在给定潜变量的条件下，类别型和有序变量之间不存在相关性。通过直接计算矩阵 $\Lambda - I$ 的主子式，而非依赖简约参数化的对角占优性质，可以确认概率分布的正性。

我们用$\mathbf{x}$ 和$\mathbf{y}$ 表示观测到的连续变量和二元虚拟变量，用$\mathbf{z}$表示连续潜变量。每个变量都是列向量，其维度分别为p_x、 q和 p_z。这里，虚拟向量$\mathbf{y}$ 的状态仅限于第 II A 节所述分类变量和有序变量允许的状态。我们将给定潜变量时观测变量的条件分布表示为互不相关的正态分布与互不相关的分类与有序分布的乘积

如下:

$$p(\mathbf{x},\mathbf{y}|\mathbf{z}) = \mathcal{N}\big(\mathbf{x} \mid \boldsymbol{\mu}_x + W(\mathbf{z}-\boldsymbol{\mu}_z), \Psi\big) \prod_{j=1}^{c} \mathrm{Cat}(\mathbf{y}_{C_j} \mid \boldsymbol{\beta}_{C_j}) \prod_{k=1}^{o} \mathrm{Ord}(\mathbf{y}_{O_k} \mid \boldsymbol{\beta}_{O_k}), \tag{57}$$

$$\boldsymbol{\beta} = \mathbf{b} + G(\mathbf{z}-\boldsymbol{\mu}_z), \quad G \equiv \Big[\mathbf{g}_1,\ \mathbf{g}_2,\ \ldots,\ \mathbf{g}_q\Big]^T, \tag{58}$$

其中，我们将类别与有序分布定义如下:

$$\mathrm{Cat}(\mathbf{y}_{C_j} \mid \boldsymbol{\beta}_{C_j}) = \frac{\exp(\mathbf{y}_{C_j}^T \boldsymbol{\beta}_{C_j})}{1+\sum_{l=1}^{q_{C_j}} \exp([\boldsymbol{\beta}_{C_j}]_l)}, \quad \mathrm{Ord}(\mathbf{y}_{O_j} \mid \boldsymbol{\beta}_{O_j}) = \frac{\exp(\mathbf{y}_{O_j}^T \boldsymbol{\beta}_{O_j})}{1+\sum_{l=1}^{q_{O_j}} \prod_{m=1}^{l} \exp([\boldsymbol{\beta}_{O_j}]_m)}. \tag{59}$$

上述条件分布由连续变量对应的参数（$\boldsymbol{\mu}_x$， Ψ， W）以及类别型和有序型虚拟变量对应的参数（$\mathbf{b}$ G）确定。 p_x维列向量$\boldsymbol{\mu}_x$ 用于参数化观测连续变量的均值，而 $p_x \times p_x$ 对角矩阵Ψ 是观测噪声的协方差矩阵。矩阵 W 是连续变量 [2]的因子载荷矩阵。 q维列向量$\mathbf{b}$ 表示类别与有序分布的偏置项，而矩阵 G 是类别型和有序型虚拟变量的因子载荷矩阵。我们将因子载荷矩阵 G 的每一行向量称为因子载荷向量$\mathbf{g}_r$。

潜变量的先验分布由正态分布混合给出，

$$p(\mathbf{z}) = \sum_{R_1 \subseteq R} \pi_{R_1}(\Sigma_z)\, \mathcal{N}(\mathbf{z} \mid \boldsymbol{\mu}_z + \Sigma_z G^T \mathbf{1}_{R_1}, \Sigma_z), \tag{60}$$

$$\pi_{R_1}(\Sigma_z) \equiv \frac{\exp\Big(\mathbf{1}_{R_1}^T \mathbf{b} + \frac{1}{2}\mathbf{1}_{R_1}^T G \Sigma_z G^T \mathbf{1}_{R_1}\Big)}{\sum_{R_1' \subseteq R} \exp\Big(\mathbf{1}_{R_1'}^T \mathbf{b} + \frac{1}{2}\mathbf{1}_{R_1'}^T G \Sigma_z G^T \mathbf{1}_{R_1'}\Big)}, \tag{61}$$

其中，求和 $\sum_{R_1 \subseteq R}$ 遍历的是类别变量和有序变量所允许的所有虚拟变量状态，而非所有可能虚拟变量的状态，并且 $\mathbf{1}_{R_1}$ 是一个 q维常量向量，其每个元素取值为**0**或**1**，

$$[\mathbf{1}_{R_1}]_s \equiv \begin{cases} 1, & \text{if } s \in R_1, \\ 0, & \text{if } s \in R_0, \end{cases} \qquad (s = 1, 2, \ldots, q). \tag{62}$$

参数（$\boldsymbol{\mu}_z$， Σ_z）是潜变量先验分布的均值和协方差。

然后，观测变量的分布为

$$p(\mathbf{x},\mathbf{y}) = \pi_{R_1}(\Sigma_z)\,\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_x + W\Sigma_z G^T\mathbf{y}, \Sigma_x), \tag{63}$$

$$\Sigma_x = \Psi + W\Sigma_z W^T. \tag{64}$$

当观测变量全部由虚拟变量构成时，观测分布就是伊辛模型 [10]。

给定观测变量时，潜变量的后验分布是正态分布：

$$p(\mathbf{z}|\mathbf{x},\mathbf{y}) = \mathcal{N}(\mathbf{z} \mid \mathbf{m}, \Sigma_{z|x}), \tag{65}$$

$$\mathbf{m} = \boldsymbol{\mu}_z + \Sigma_{z|x}\big\{W^T\Psi^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) + G^T\mathbf{y}\big\}, \tag{66}$$

$$\Sigma_{z|x} = \big(\Sigma_z^{-1} + W^T\Psi^{-1}W\big)^{-1}. \tag{67}$$

我们将上述表达式中的 $\mathbf{m}$ 称为因子得分。

A. 双标图分析

在因子分析的双标图中，我们首先将因子得分（式 (66)）绘制为散点图。不失一般性，设 $\boldsymbol{\mu}_z = \mathbf{0}$ 和 $\Sigma_z = I$。由于类别变量和序数变量由虚拟变量组合表示，因此用因子载荷向量的组合而非单个因子载荷向量来表示双标图中以箭头描绘的因子载荷向量是恰当的。我们将类别变量和序数变量的组合因子载荷向量定义如下：

$$\mathbf{g}_l^{(C_j)} = -\frac{1}{q_{C_j}+1}\mathbf{g}_{C_j} + \mathbf{g}_{\min(C_j)+l-1}, \quad (l = 1, 2, \ldots, q_{C_j}), \tag{68}$$

$$\mathbf{g}_0^{(C_j)} = -\frac{1}{q_{C_j}+1}\mathbf{g}^{C_j}, \tag{69}$$

$$\mathbf{g}_l^{(O_j)} = -\frac{1}{2}\mathbf{g}_{O_j} + \sum_{m=1}^{l}\mathbf{g}_{\min(O_j)+m-1}, \quad (l = 1, 2, \ldots, q_{O_j}), \tag{70}$$

$$\mathbf{g}_0^{(O_j)} = -\frac{1}{2}\mathbf{g}^{O_j}. \tag{71}$$

新的组合因子载荷向量，$\mathbf{g}^{(C_j)\ 0}$ 和$\mathbf{g}^{(O_j)\ 0}$，分别对应基准类别和基准水平，其中属于特定分类或有序变量的所有虚拟变量

取值为零，并满足

$$\mathbf{g}_0^{(C_j)} = -\sum_{l=1}^{q_{C_j}} \mathbf{g}_l^{(C_j)}, \qquad \mathbf{g}_0^{(O_j)} = -\mathbf{g}_{q_{O_j}}^{(O_j)}. \tag{72}$$

组合因子载荷向量与因子得分的内积表示分类与有序分布参数的相对大小。也就是说，内积值较大意味着，在相应的因子得分处，对应的分类或序数特征比均值更有可能出现。

我们施加了范数约束，使所有组合因子载荷向量具有相同的范数，以避免因子分析中最大似然估计的不适当解，即 Heywood 情形；该方法已在参考文献 [7]中提出。此外，与前一篇论文 [7], 一样，我们固定了潜在空间的旋转自由度，使矩阵GG^T 对角化。我们通过 GG^T的特征值定义潜在空间（即主成分轴）的贡献率，并按照贡献率降序排列主成分轴。

我们对上一节分析的读者数据进行了因子分析。模型参数通过最大似然估计进行估计。潜在维度的数量通过贝叶斯信息准则 [11]进行选择。最终选定的潜在维度数量为两个。图2展示了因子分析的双标图。图中显示了第一和第二主成分轴（PC1、PC2），即潜在维度；轴标签中的百分比表示贡献率。当压缩为二维时，序数变量的不等间隔被可视化为量角器的角度。我们可以看到，第一主成分轴可以解释为代表受教育知识，并且与女性杂志的阅读行为相关。

V. 结论

我们提出了一种方法，在格拉斯曼分布框架下结合类别型和有序变量的虚拟变量编码，来表示多元分类与序数变量之间的相关性。所提出的分布能够使类别型和有序型虚拟变量中不理想的共现概率精确为零。此外，与量化方法不同，所提出的方法在处理类别型和有序变量时保留了其尺度性质，即名义尺度和顺序尺度。

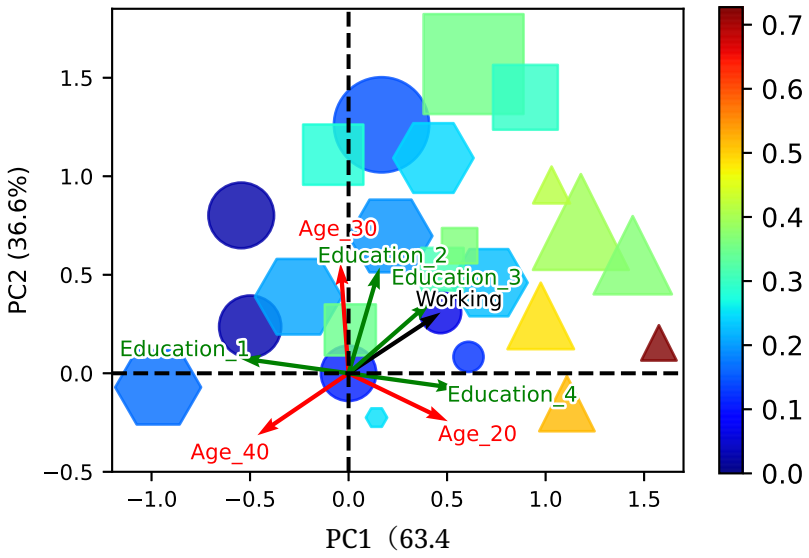


图2. 读者数据的因子分析双标图。散点图中的点符号根据固定读者阅读某一女性杂志的百分比所处的四分位区间而变化：圆形表示百分比低于第一四分位数的数据，六边形表示介于第一和第二四分位数之间的数据，正方形表示介于第二和第三四分位数之间的数据，三角形表示高于第三四分位数的数据。这些标记根据百分比由低到高从蓝色渐变为红色。散点图中点符号的面积与对应点符号的数据点数量成正比。

作为所提出分布的一项应用，我们针对类别型和有序变量开发了一种因子分析方法。我们的因子分析双标图可作为对应分析的替代方法 [12]。

从以往研究 [6, 7] 到本文的一系列研究，已经为四种统计数据类型中的三种主要类型——比率尺度、名义尺度和顺序尺度 [13]——构建了描述线性相关的概率分布。这些分布使我们能够为连续变量和定性变量的混合形式构建概率生成模型。我们预计，这些研究成果将为定性数据的定量分析开辟新的方向，并为构建更先进的统计学习方法奠定基础。

附录 A：反向格拉斯曼分布

在本附录中，为避免可能的混淆，我们总结了这样一种概率分布：其中，虚拟变量的影响与以往研究 [6, 7], 中的影响相反。

我们分别用列向量 $\mathbf{x}$ 和$\mathbf{y}$表示 p 维连续变量和 q 维虚拟变量。虚拟向量 $\mathbf{y}$ 是一个比特向量，其中每个元素取值为 0 或 1。我们的分布的模型参数包括多元正态分布的均值和协方差参数（$\boldsymbol{\mu}$, Σ）、格拉斯曼分布 Λ [6] 的 $q \times q$ 矩阵，以及表示连续变量与二元变量之间交互作用的 $q \times p$ 矩阵 G。矩阵 $\Lambda - I$ 必须是一个 P_0 矩阵 [8], ，其中 I 是单位矩阵。矩阵 G、$[G]_{sj}$、（$s = 1$、2、.....、q、以及 $j = 1$、2、.....、p）的每个元素也表示为一个 p 维列向量 $\mathbf{g}_s$ ，即

$$G \equiv \Big[\mathbf{g}_1,\ \mathbf{g}_2,\ \ldots,\ \mathbf{g}_q\Big]^T, \quad [G]_{sj} = [\mathbf{g}_s^T]_j, \tag{A1}$$

其中，T 表示矩阵转置。我们分别将连续变量和二元变量的全部索引集合记为 $I \equiv \{1, 2, \ldots, p\}$ 和 $R \equiv \{1, 2, \ldots, q\}$。二元变量的索引标签分为带下标 1 和 0 的两部分，分别对应取值为 1 和 0 的变量。例如，二元变量的索引标签 R 被划分为一个子集 $R_1 \subseteq R$ 及其集合差 $R_0 = R \backslash R_1$。我们用 q 维常向量 $\mathbf{1}_{R_1}$ 表示其每个元素取值为 0 或 1，

$$[\mathbf{1}_{R_1}]_s \equiv \begin{cases} 1, & \text{if } s \in R_1 \\ 0, & \text{if } s \in R_0 \end{cases}, \quad (s = 1, 2, \ldots, q). \tag{A2}$$

于是，所提出的联合分布表示为

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} = \mathbf{1}_{R_1}) =& \pi_{R_1}(\Sigma)\, \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu} + \Sigma G^T \mathbf{1}_{R_1}, \Sigma), \\ \equiv& \pi_{R_1}(\Sigma) \frac{1}{(2\pi)^{p/2} \det \Sigma^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \Sigma G^T \mathbf{1}_{R_1})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu} - \Sigma G^T \mathbf{1}_{R_1})\right\}, \end{aligned} \tag{A3}$$

$$\pi_{R_1}(\Sigma) \equiv \frac{\det(\Lambda_{R_1 R_1} - I) \exp\left(\frac{1}{2} \mathbf{1}_{R_1}^T G \Sigma G^T \mathbf{1}_{R_1}\right)}{\sum_{R_1' \subseteq R} \det(\Lambda_{R_1' R_1'} - I) \exp\left(\frac{1}{2} \mathbf{1}_{R_1'}^T G \Sigma G^T \mathbf{1}_{R_1'}\right)}, \tag{A4}$$

其中，$\Lambda_{R_1 R_1}$ 是 Λ的一个子矩阵，求和$\sum_{R_1' \subseteq R}$ 遍历二元变量的所有可能状态。该分布的配分函数（即归一化常数）为

无法通过解析方法求得，因此必须对二元变量的所有可能状态求和，才能计算配分函数。正如我们将在下文看到的，系数 $\pi_{R_1}(\Sigma)$ 对应于协方差相同的高斯分布混合中的混合权重。也就是说，上述联合分布对应于 2^q 正态分布混合中的一个正态分布。

为表示边缘分布和条件分布，我们首先定义指标记法。我们将完整指标集合的某个子集的索引标签记为 $J \subseteq I$。随后，由索引子集 J 构成的子向量表示为 $\mathbf{x}_J$。我们将连续变量和二元变量的完整指标集合分别划分为三个子集部分： $I = (J, L, K)$ 和 $R = (S, U, T)$，其中引入索引标签 L 和 U 以处理缺失值。这些指标集合中的元素数量分别表示为 p_J、p_L、p_K和 q_S、q_U、q_T，显然它们满足 $p_J + p_L + p_K = p$和 $q_S + q_U + q_T = q$。于是，向量 $\mathbf{x}$ 和 $\mathbf{y}$ 可分别划分为如下子向量：$\mathbf{x} = \mathbf{x}_I = (\mathbf{x}_J, \mathbf{x}_L, \mathbf{x}_K)$ 和 $\mathbf{y} = \mathbf{y}_R = (\mathbf{y}_S, \mathbf{y}_U, \mathbf{y}_T)$。同样，二元变量的索引标签还可根据下标 $\mathbf{1}$ 和 $\mathbf{0}$ 进一步划分为两部分，分别对应取值为 $\mathbf{1}$ 和 $\mathbf{0}$ 的变量。例如，二元变量的索引标签 $S \subseteq R$ 可划分为子集 $S_1 \subseteq S$ 及其集合差 $S_0 = S \setminus S_1$，其中子向量 $\mathbf{y}_{S_1}$ 和 $\mathbf{y}_{S_0}$ 分别取值为 $\mathbf{y}_{S_1} = \mathbf{1}$ 和 $\mathbf{y}_{S_0} = \mathbf{0}$。索引标签 J和 K 的并集记为 $J + K \equiv J \cup K$。利用上述指标记法，边缘分布表示为

$$p(\mathbf{x}_K = \mathbf{x}_{I\setminus(J+L)}, \mathbf{y}_T = \mathbf{y}_{R\setminus(S+U)}) = \sum_{S_1+U_1 \subseteq R\setminus T} \pi_{R_1}(\Sigma)\, \mathcal{N}(\mathbf{x}_K \mid \boldsymbol{\mu}_K + \Sigma_{KI} G^T \mathbf{1}_{R_1}, \Sigma_{KK}). \tag{A5}$$

特别是，当所有二元变量都被边缘化时，边缘分布恰好是一个 2^q 协方差相同的高斯分布混合，其混合权重由 $\pi_{R_1}(\Sigma)$ 给出，且正态分布的均值偏移了 $\Sigma G^T \mathbf{1}_{R_1}$。另一方面，当所有连续变量都被边缘化时，边缘分布不再具有与格拉斯曼分布相同的形式。

T缺失值条件下，$\mathbf{x}_L$ 和 $\mathbf{y}_U$ 的条件分布为 by

$$p(\mathbf{x}_J, \mathbf{y}_S|\mathbf{x}_K, \mathbf{y}_T) = \frac{\sum_{U_1 \subseteq R\setminus(S+T)} \pi_{R_1}\big(\Sigma_{(J+L)|K}\big) \exp\Big\{\mathbf{1}_{R_1}^T G \Sigma_{IK} \Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K)\Big\}}{\sum_{S_1'+U_1' \subseteq R\setminus T} \pi_{R_1'}\big(\Sigma_{(J+L)|K}\big) \exp\Big\{\mathbf{1}_{R_1'}^T G \Sigma_{IK} \Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K)\Big\}}$$
$$\mathcal{N}\big(\mathbf{x}_J \mid \boldsymbol{\mu}_J + \Sigma_{JK}\Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K) + (\Sigma_{J(J+L)} - \Sigma_{JK}\Sigma_{KK}^{-1}\Sigma_{K(J+L)}) G_{(J+L)R}^T \mathbf{1}_{R_1}, \Sigma_{J|K}\big), \tag{A6}$$

其中， Σ_{KK}^{-1} 表示子矩阵 Σ_{KK}的逆矩阵，矩阵 $\Sigma_{J|K} \equiv \Sigma_{JJ} - \Sigma_{JK}\Sigma_{KK}^{-1}\Sigma_{KJ}$为舒尔补，混合权重的定义如前所述，

$$\pi_{R_1}\big(\Sigma_{(J+L)|K}\big) \equiv \frac{\det(\Lambda_{R_1R_1} - I) \exp\Big\{\frac{1}{2}\mathbf{1}_{R_1}^T G_{R(J+L)} \Sigma_{(J+L)|K} G_{(J+L)R}^T \mathbf{1}_{R_1}\Big\}}{\sum_{R_1' \subseteq R} \det(\Lambda_{R_1'R_1'} - I) \exp\Big\{\frac{1}{2}\mathbf{1}_{R_1'}^T G_{R(J+L)} \Sigma_{(J+L)|K} G_{(J+L)R}^T \mathbf{1}_{R_1'}\Big\}}. \tag{A7}$$

当不存在缺失值时，条件分布可以更简洁地表示为:

$$p(\mathbf{x}_J, \mathbf{y}_S|\mathbf{x}_K = \mathbf{x}_{I\setminus J}, \mathbf{y}_T = \mathbf{y}_{R\setminus S}) = \frac{\pi_{R_1}(\Sigma_{J|K}) \exp\Big\{\mathbf{1}_{S_1}^T G \Sigma_{IK} \Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K)\Big\}}{\sum_{S_1' \subseteq R\setminus T} \pi_{R_1'}(\Sigma_{J|K}) \exp\Big\{\mathbf{1}_{S_1'}^T G \Sigma_{IK} \Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K)\Big\}}$$
$$\mathcal{N}\big(\mathbf{x}_J \mid \boldsymbol{\mu}_J + \Sigma_{JK}\Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K) + \Sigma_{J|K} G_{JR}^T \mathbf{1}_{R_1}, \Sigma_{J|K}\big). \tag{A8}$$

特别地，当观测变量全部由二元变量构成时，条件分布可表示为正态分布，

$$p(\mathbf{x}_J|\mathbf{x}_K = \mathbf{x}_{I\setminus J}, \mathbf{y}_R) = \mathcal{N}\left(\mathbf{x}_J \mid \boldsymbol{\mu}_J + \Sigma_{JK}\Sigma_{KK}^{-1}(\mathbf{x}_K - \boldsymbol{\mu}_K) + \Sigma_{J|K} G^T \mathbf{1}_{R_1}, \Sigma_{J|K}\right), \tag{A9}$$

此时，分布的均值会根据所条件化的二元变量的取值而偏移。另一方面，当观测变量全部由连续变量构成时，条件分布可表示为格拉斯曼分布:

$$p(\mathbf{y}_S|\mathbf{x}_I, \mathbf{y}_T = \mathbf{y}_{R\setminus S}) = \mathcal{G}^{(1)}\big(\mathbf{y}_S \mid I + (\Lambda - I)_{S|T_1} \operatorname{Exp}\big\{G_{SI}(\mathbf{x}_I - \boldsymbol{\mu}_I)\big\}\big),$$
$$\equiv \frac{\det\big[(\Lambda - I)_{S_1|T_1} \operatorname{Exp}\big\{G_{S_1I}(\mathbf{x}_I - \boldsymbol{\mu}_I)\big\}\big]}{\det\big[I + (\Lambda - I)_{S|T_1} \operatorname{Exp}\big\{G_{SI}(\mathbf{x}_I - \boldsymbol{\mu}_I)\big\}\big]}, \tag{A10}$$
$$(\Lambda - I)_{S|T_1} \equiv \Lambda_{SS} - I - \Lambda_{ST_1}(\Lambda_{T_1T_1} - I)^{-1}\Lambda_{T_1S}, \tag{A11}$$

其中

$$\operatorname{Exp}\big\{G_{SI}(\mathbf{x}_I - \boldsymbol{\mu}_I)\big\} \equiv E^{G_{SI}(\mathbf{x}_I - \boldsymbol{\mu}_I)} \equiv \operatorname{diag}\big[\exp\big\{\mathbf{g}_s^T(\mathbf{x}_I - \boldsymbol{\mu}_I)\big\}\big], \quad s \in S \tag{A12}$$

是一个对角元素非负的对角矩阵。

参考文献

---

[1] R. Likert, 《心理学档案》**140**，5（1932）。[2] K. P. Murphy, 《机器学习：概率视角》（MIT出版社，2012）。[3] M. E. Tipping，载于第*11*届神经信息处理系统国际会议论文集，NIPS'98（MIT出版社，美国马萨诸塞州剑桥，1998），第592–598页。[4] M. Collins、S. Dasgupta 和 R. E. Schapire，载于神经信息处理系统进展，第14卷，T. Dietterich、S. Becker 和 Z. Ghahramani 编（MIT出版社，2001）。[5] S. Mohamed、Z. Ghahramani 和 K. A. Heller，载于神经信息处理系统进展，第21卷，D. Koller、D. Schuurmans、Y. Bengio 和 L. Bottou 编（Curran Associates, Inc.，2008）。[6] T. Arai, 《物理评论E》**103**，062104（2021）。[7] T. Arai, （2022），10.48550/ARXIV.2209.12147。[8] M. J. Tsatsomeros, 《亚洲信息科学生活：国际期刊》**1**，115（2002）。[9]G. T. Gunther Schauberger，"catdata"，CRAN软件包（2020），网址：https://CRAN.R- project.org/package=catdata（访问日期：2022.8.26）。[10] E. Ising, 《Zeitschrift für Physik》**31**，253（1925）。[11] G. Schwarz, 《统计年鉴》**6**，461（1978）。[12] H. O. Hirschfeld，剑桥哲学学会数学论文集**31**，520（1935）。[13] S. S. Stevens, 《科学》**103**，677（1946）。