arXiv:2303.04286v1 [stat.ME] 7 Mar 2023

特征矩阵的充分降维

Chanwoo Lee

统计系，威斯康星大学麦迪逊分校
chanwoo.lee@wisc.edu

摘要

我们研究特征矩阵的充分降维问题，这一问题常见于传感器网络定位、脑神经影像和脑电图分析。一般而言，特征矩阵既可以按行解释，也可以按列解释，并包含结构信息，而朴素的向量化方法可能会丢失这些信息。为此，我们提出一种称为主支撑矩阵机（PSMM）的矩阵充分降维方法。PSMM 将响应变量划分为多个切片，从而将充分降维问题转化为一系列分类问题。该方法通过寻找法矩阵秩为 1、能够最优地区分切片响应的超平面，有效利用矩阵结构。此外，我们还将该方法推广到高阶张量情形。数值分析表明，PSMM 优于现有方法，并且在真实数据应用中具有很强的可解释性。

关键词： 充分维度约简，支撑矩阵机，维度折叠，主支撑向量机

1 引言

矩阵值数据集在现代数据科学应用中无处不在。例如，脑电图（EEG）数据收集了两个组别中122名受试者的数据：酒精依赖组和对照组。每名受试者都接受了刺激，头皮上安装了64个电极，记录了256个时间点的电压值。因此，每个抽样单元由一个 256×64 带有组标签的矩阵构成。了解酒精依赖与不同时间和通道上的电压模式之间的关系具有科学意义。另一个例子是MRN-114人脑连接数据。该数据包含114名受试者及其全量表智商（FSIQ）得分。对于每名受试者，研究人员基于Desikan图谱收集了68个脑区之间的二元连接矩阵（Desikan等人，2006）。从这一矩阵值数据集中学习，可以深入了解智力与脑连接之间的关联。

设 $\boldsymbol{X} \in \mathbb{R}^{d_1 \times d_2}$ 为矩阵预测变量， $Y \in \mathbb{R}$ 为响应变量。我们希望在不丢失 $\boldsymbol{X}$ 与 Y之间回归关系的前提下，降低矩阵 $\boldsymbol{X}$ 的维度。一种简单的方法是将特征矩阵向量化，并应用经典降维方法来估计矩阵 $\boldsymbol{M} \in \mathbb{R}^{d_1 d_2 \times r}$ ，使得$r < d_1 d_2$

$$Y \perp\!\!\!\perp \mathrm{vec}(\boldsymbol{X}) | \boldsymbol{M}^T \mathrm{vec}(\boldsymbol{X}), \tag{1.1}$$

其中 vec： $\mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 d_2}$ 是将矩阵转换为列向量的线性变换。这个经典的充分降维问题受到了广泛关注，已有许多方法被提出并得到研究（Li，1991；Duan和Li，1991；Cook和Weisberg，1991；Cook和Ni，2005；Li 等人，2005；Li和Wang，2007；Artemiouand Dong，2016）。

然而，矩阵并非只是带有额外索引的向量；相反，它们具有结构信息，而简单的向量化方法无法利用这些信息。例如，在脑电图数据中，矩阵 $\boldsymbol{X}$ 的每一行和每一列都对应一个特定的电极通道和时间点。若将这个特征矩阵向量化，这种有意义的解释就会丢失。类似地，在 MRN-114人脑连接数据中，脑网络通常表示为对称邻接矩阵 $\boldsymbol{X}$，其中的值表示纤维连接是否存在。将这些矩阵转换为向量会破坏其对称性，导致其中包含的信息无法得到充分利用。此外，如果通过堆叠特征矩阵的列或行将其转换为向量，所得向量的维度会非常高，从而可能导致维数灾难。将 $\boldsymbol{X}$ 保留为矩阵，可以减少参数数量，并提高估计的准确性。

为利用矩阵结构，我们考虑以下目标函数，以求得两个矩阵 $\boldsymbol{U} \in \mathbb{R}^{d_1 \times r_1}$ $(r_1 < d_1)$ 和 $\boldsymbol{V} \in \mathbb{R}^{d_2 \times d_2}$ $(r_2 < d_2)$，使得

$$Y \perp\!\!\!\perp \boldsymbol{X} | \boldsymbol{U}^T \boldsymbol{X} \boldsymbol{V}. \tag{1.2}$$

然后，我们保留特征矩阵 $\boldsymbol{X} \in \mathbb{R}^{d_1 \times d_2}$ 中的所有信息，以预测响应变量Y，同时将其维度降至 $\boldsymbol{U}^T \boldsymbol{X} \boldsymbol{V} \in \mathbb{R}^{r_1 \times r_2}$。与此同时，我们仍保留矩阵结构信息，包括按行和按列解释。

请注意，可识别参数是由矩阵张成的子空间$\boldsymbol{U}$ 和 $\boldsymbol{V}$，即 span($\boldsymbol{U}$) 和 span($\boldsymbol{V}$)。事实上，(1.2) 等价于

$$Y \perp\!\!\!\perp \boldsymbol{X} | (\boldsymbol{U}\boldsymbol{A})^T \boldsymbol{X} (\boldsymbol{V}\boldsymbol{B}),$$

对于任意非奇异矩阵 $\boldsymbol{A} \in \mathbb{R}^{r_1 \times r_1}$ 和 $\boldsymbol{B} \in \mathbb{R}^{r_2 \times r_2}$。因此，我们的目标是估

计 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的列空间，而不是 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 本身。我们将由 $\boldsymbol{U}$ （以及 $\boldsymbol{V}$）所诱导且满足（1.2）的子空间称为行（和列）降维子空间。这一概念最早由Li 等人（2010）提出，称为左、右维度折叠子空间。我们仿照向量情形中中心子空间的定义，定义中心行子空间和中心列子空间。在向量情形中，众所周知，两个降维子空间的交本身也是一个降维子空间（Chiaromonte 和 Cook，2002；Yin 等人，2008）。类似地，在矩阵情形中，关于 $Y|\boldsymbol{X}$ 的行降维子空间的交，在温和条件下仍是一个行降维子空间（Li 等人，2010）。列降维子空间也有同样的论证。因此，我们按如下方式定义中心行子空间和中心列子空间。

定义 1（矩阵中心子空间）。 定义中心行子空间和中心列子空间如下

$$S^r_{Y|\boldsymbol{X}} = \bigcap_{\boldsymbol{U}:\, Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}} \operatorname{span}(\boldsymbol{U}) \quad \text{and} \quad S^c_{Y|\boldsymbol{X}} = \bigcap_{\boldsymbol{V}:\, Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}} \operatorname{span}(\boldsymbol{V}).$$

子空间 $S_{Y|\boldsymbol{X}} = S^r_{Y|\boldsymbol{X}} \times S^c_{Y|\boldsymbol{X}}$ 称为矩阵中心子空间。

我们还可以将中心子空间改写为

$$S_{Y|\boldsymbol{X}} = \bigcap_{(\boldsymbol{U},\boldsymbol{V}):\, Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}} \operatorname{span}(\boldsymbol{U}) \times \operatorname{span}(\boldsymbol{V}).$$

Li 等人Li 等人（2010）；Ding 和 Cook （2015）与我们定义的中心子空间等价，只需将笛卡尔积 $\times$替换为克罗内克积 $\otimes$。这里，为了使阐述更简洁，并便于推广到高阶张量情形，我们采用笛卡尔积。高阶张量的中心子空间的推广见第 4节。

本文旨在估计矩阵的中心子空间，并提出了主支撑矩阵机（PSMM）。

1.1 相关工作与我们的贡献

我们的研究与现有的若干研究方向密切相关，但也存在明显差异。本节回顾相关文献，并阐述我们的贡献。

主支撑向量机 我们的 PSMM 与以下文献提出的主支持向量机（PSVM）密切相关：Li 等人（2011）；Artemiou and Dong（2016）。PSVM 针对向量情形，考虑式（1.1）中的充分降维。PSVM 的主要思路是根据响应变量的取值，将特征向量划分为若干切片，并利用支持向量机求得能够最佳分隔这些切片的超平面。通过主成分分析汇总这些超平面，可得到向量情形下中心子空间的一致估计量。然而，PSVM 仅适用于向量型预测变量；将矩阵型预测变量向量化会损失结构信息，并导致维数过高。相比之下，我们的 PSMM 为矩阵预测变量提供了一种高效的充分降维方法，并能成功保留结构信息。数值研究表明，基于矩阵的方法具有明显优势。

支持矩阵机 支持矩阵机（SMM）已成功应用于图像分类、视觉识别和脑电图数据分析（Pirsiavash 等人，2009；Luo 等人，2015）。SMM 是针对矩阵预测变量的分类问题提出并发展的。它考虑以下公式——

……，将支持向量机扩展到矩阵情形：

$$\min_{\boldsymbol{B},b} \|\boldsymbol{B}\|_F^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\left\{1 - Y_i(\langle \boldsymbol{B}, \boldsymbol{X}_i - \bar{\boldsymbol{X}}\rangle + b)\right\}_+, \tag{1.3}$$

其中 $\{x\}_+ = \max(x,\ 0)$，$\boldsymbol{B} \in \mathbb{R}^{d_1 \times d_2}$ 是一个具有低秩性的系数矩阵，$b \in \mathbb{R}$ 是截距，λ 是正的惩罚参数。通过对系数正规矩阵施加低秩结构，SMM 利用了特征矩阵的结构信息。我们借鉴这一思路，并将其应用于充分降维领域。我们将充分降维问题转化为一系列分类问题，并在其中使用经过改进的 SMM 技术。我们证明，PSMM 能够有效估计矩阵中心子空间，同时兼具 SMM 方法的优势，例如考虑矩阵结构、抵抗离群值的稳健性，以及在高维情形下的效率。

维度折叠Li 等人（2010）；Ding和Cook （2015）提出了中心维度折叠子空间，该子空间等同于定义1中所定义的矩阵中心子空间。他们还提出了估计中心子空间的方法，将现有的基于逆回归的方法推广到矩阵和高阶张量情形。这类方法包括切片逆回归（SIR）（Li，1991）、切片平均方差估计量（SAVE）（Cook和Weisberg，1991）和方向回归（降维）（Li和Wang，2007）。然而，正如Li 等人（2011）所指出的，由于长度较短，这类方法往往会降低数据中心附近切片均值的权重。这一特性常使这些方法不够准确，因为众所周知，回归曲面在数据中心处的估计效果良好。相比之下，PSMM 能够找到系数法矩阵，根据切片响应实现数据点的最优区分。这种方法可以适当利用中心附近的数据点。我们证明，PSMM 的准确性确实优于基于逆

第5节中的基于回归的方法。

1.2 符号与结构

我们使用简写 $[n]$ 表示 $\{1,\ldots,n\}$ 的 $n \in \mathbb{N}_+$。对于任意两个矩阵 $\boldsymbol{A},\boldsymbol{B} \in \mathbb{R}^{d_1\times d_2}$，两个矩阵的内积定义为 $\langle \boldsymbol{A},\boldsymbol{B}\rangle = \sum_{(i,j)\in[d_1]\times[d_2]} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$。对于矩阵 $\boldsymbol{A} \in \mathbb{R}^{d_1\times d_2}$，我们用 $\lambda_i(\boldsymbol{A})$ 表示 $\boldsymbol{A}$ 的第 i 大特征值，并用 $\|\boldsymbol{A}\|_F = \sqrt{\sum_{(i_1,i_2)\in[d_1\times[d_2]} \boldsymbol{A}^2_{i_1i_2}}$ 表示其 Frobenius 范数。设 $\mathcal{A} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ 为一个阶数为 K 的（$d_1,\ldots,d_K$）维张量，且 $\mathcal{A}_{i_1,\ldots,i_K}$ 为张量中由（$i_1,\ldots,i_K$）$\in [d_1]\times\cdots\times[d_K]$ 索引的元素。我们将张量 $\mathcal{A}$ 的 Frobenius 范数定义为 $\|\mathcal{A}\|_F = \sqrt{\sum_{(i_1,\ldots,i_K)\in[d_1]\times\cdots\times[d_K]} \mathcal{A}^2_{i_1,\ldots,i_K}}$。张量 $\mathcal{C} \in \mathbb{R}^{r_1,\ldots,r_K}$ 与矩阵 $\boldsymbol{U}_k \in \mathbb{R}^{d_k\times r_k}$, $k \in [K]$ 的多线性乘法定义为

$$(\mathcal{C}\times_1 \boldsymbol{U}_1 \times\cdots\times_K \boldsymbol{U}_K)_{i_1,\ldots,i_K} = \sum_{j_1=1}^{r_1}\cdots\sum_{j_K=1}^{r_K} \mathcal{C}_{j_1,\ldots,j_d}(\boldsymbol{U}_1)_{i_1j_1}\cdots(\boldsymbol{U}_K)_{i_Kj_K},$$

这会得到一个阶数为 K（d_1，……，d_K）的张量。

本文其余部分安排如下。第2 节介绍总体层面的 PSMM 目标函数，并构造估计量的无偏性。随后，我们在第 3 节介绍样本层面的矩阵充分降维估计程序。在第 4 节中，我们将矩阵情形的所有结果推广到高阶张量情形。第 5 节呈现合成数据和真实数据分析。最后，我们在第6节进行讨论并总结全文。

2 总体层面的主支撑矩阵机

本节介绍总体层面的 PSMM 目标函数，并阐释 PSMM 用于矩阵充分降维的直观思路。

我们首先考虑特征矩阵的二分类问题。目前，我们假设响应变量 Y 取值为 -1 或 1。我们介绍秩-1 支持矩阵机（SMM），其样本为 $\{(Y_i,\ \boldsymbol{X}_i)\}_{i=1}^n$。将秩-1 系数矩阵 $\boldsymbol{B}=\boldsymbol{u}\boldsymbol{v}^T$ 代入方程（1.3）中的 SMM，得到秩-1 SMM：

$$\min_{(\boldsymbol{u},\boldsymbol{v},t)\in\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}} \quad (\boldsymbol{u}^T\boldsymbol{u})(\boldsymbol{v}^T\boldsymbol{v})+\frac{\lambda}{n}\sum_{i=1}^{n}\left\{1-Y_i[\boldsymbol{u}^T(\boldsymbol{X}_i-\bar{\boldsymbol{X}})\boldsymbol{v}-t]\right\}_+. \tag{2.4}$$

该 SMM 目标函数是对支持向量机的扩展，而秩-1约束有助于利用矩阵预测变量的结构信息。方程（2.4）的解记为（$\boldsymbol{u}^*,\ \boldsymbol{v}^*,\ t^*$），它定义了最优超平面 $\{\boldsymbol{X}:\ (\boldsymbol{u}^*)^T\boldsymbol{X}\boldsymbol{v}^*=t^*\}$，该超平面将两个空间 $\{\boldsymbol{X}_i:\ Y_i=1\}$ 和 $\{\boldsymbol{X}_i:\ Y_i=-1\}$分隔开。

现在我们考虑矩阵充分降维问题，其中响应变量 Y 可以是连续变量。令 Ω_Y 为 Y的支撑集。令 A_1 和 $\tilde{A_2}$为 Ω_Y的任意不相交子集。定义离散随机变量 $\tilde{Y}$，使得

$$\tilde{Y}=\mathbb{1}\{Y\in A_1)-\mathbb{1}\{Y\in A_2\}.$$

我们为矩阵充分降维提出如下总体层面的目标函数：

$$L(\boldsymbol{u},\boldsymbol{v},t)=\mathrm{Var}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v})+\lambda\mathbb{E}\left\{1-\tilde{Y}(\boldsymbol{u}^T(\boldsymbol{X}-\mathbb{E}(\boldsymbol{X}))\boldsymbol{v}-t)\right\}_+. \tag{2.5}$$

与方程（2.4）中的秩为 1 的 SMM 相比，我们考虑$\boldsymbol{X}$的方差因子。若 Var $[\mathrm{vec}(\boldsymbol{X})]=I_{d_1+d_2}$，则方程（2.5）中的目标函数将简化为方程（2.4）的总体版本。考虑该方差可确保使（2.5）最小化的估计量对中心子空间无偏，正如下述定理所示。

定理*1.* 假设 $\mathbb{E}(\boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V})$ 是关于 $\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}$ 的双线性函数，其中 $\boldsymbol{U}$和 $\boldsymbol{V}$是式 *(1.2)* 中定义的矩阵。若 $(\boldsymbol{u}^*, \boldsymbol{v}^*, t^*)$ 在所有满足 $\in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}$ 的 $(\boldsymbol{u}, \boldsymbol{v}, t)$ 中使目标函数 *(2.5)* 取得最小值，则 $(\boldsymbol{u}^*, \boldsymbol{v}^*) \in S_{Y|\boldsymbol{X}}$.

证明. 不失一般性，假设 $\mathbb{E}(\boldsymbol{X}) = 0_{d_1\times d_2}$。注意到

$$\mathbb{E}\left[1-\tilde{Y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}-t)\right]_+ = \mathbb{E}\left[\mathbb{E}\left[(1-\tilde{Y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}-t))_+|Y,\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right]\right]$$

根据詹森不等式，我们有

$$\begin{aligned}
&\mathbb{E}\left[(1-\tilde{Y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}-t))^+|Y,\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right] \\
&\geq \mathbb{E}\left[(1-\tilde{Y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}-t))|Y,\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right]_+ \\
&= \left[1-\tilde{Y}\left[\mathbb{E}((\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v})|Y,\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V})-t\right]\right]_+,
\end{aligned}$$

其中，第一个等号由 $Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}$ 得出。因此，我们有

$$\begin{aligned}
\mathbb{E}\left[1-\tilde{Y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}-t)\right]_+ &\geq \mathbb{E}\left[1-\tilde{Y}\left[\mathbb{E}((\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v})|Y,\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V})-t\right]\right]_+ \\
&= \mathbb{E}\left[1-\tilde{Y}\left[(\boldsymbol{U}\eta_r)^T\boldsymbol{X}(\boldsymbol{V}\eta_c)-t\right]\right]_+, \qquad (2.6)
\end{aligned}$$

对于某个 $\eta_r \in \mathbb{R}^{r_1}, \eta_c \in \mathbb{R}^{r_2}$。我们总能找到这样的 η_r, η_c，因为 $\mathbb{E}(\boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V})$ 是一个

$\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}$的双线性函数。此外，注意到

$$\begin{aligned}
&\mathrm{Var}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}) \\
&\quad = \mathrm{Var}\left[\mathbb{E}\left(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right)\right] \\
&\quad \geq \mathrm{Var}\left[\mathbb{E}\left(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right)\right] \\
&\quad = \mathrm{Var}\left[(\boldsymbol{U}\eta_r)^T\boldsymbol{X}(\boldsymbol{V}\eta_c)\right], \qquad (2.7)
\end{aligned}$$

其中，最后一个等式来自条件期望的双线性。将 (2.6) 和 (2.7) 代入 (2.5) 中的目标函数，得到：

$$\begin{aligned}
L(\boldsymbol{u},\boldsymbol{v},t) &\geq \mathrm{Var}\left[(\boldsymbol{U}\eta_r)^T\boldsymbol{X}(\boldsymbol{V}\eta_c)\right] + \lambda\mathbb{E}\left[1-\tilde{Y}\left[(\boldsymbol{U}\eta_r)^T\boldsymbol{X}(\boldsymbol{V}\eta_c)-t\right]\right]_+ \\
&\geq L(\boldsymbol{U}\eta_r,\boldsymbol{V}\eta_c,t). \qquad (2.8)
\end{aligned}$$

假设$(\boldsymbol{u},\boldsymbol{v}) \notin S_{Y|\boldsymbol{X}}$，则 $\mathrm{Var}\left(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}\right) > 0$，这意味着(2.7)中的不等式是严格的。因此，(2.8)中的不等式是严格的。(2.8)中的这个严格不等式证明，除非$(\boldsymbol{u},\boldsymbol{v})$是 $L(\boldsymbol{u},\boldsymbol{v},t)$ 的最小化元，否则$(\boldsymbol{u},\boldsymbol{v}) \in S_{Y|\boldsymbol{X}}$。

□

关于 $\mathbb{E}(\boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V})$ 的双线性条件，是对线性条件的推广；在线性条件下，向量情形的充分降维文献中通常会采用这一广为人知的假设(Li and Dong, 2009; Li 等人, 2011;Artemiou and Dong,2016)。

定理1 表明，我们可以通过最小化 $S_{Y|\boldsymbol{X}}$ 一系列˜方程（2.5）中具有不同 Ys 的目标函数来估计中心子空间。下一节将基于这一思路，在样本层面提出 PSMM 的估计程序。

3 矩阵充分降维的估计程序

我们首先介绍 flip-flop 算法，以根据独立同分布样本 $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ 估计均值和协方差矩阵。随后，我们介绍 PSMM 程序，以在样本层面估计矩阵中心子空间。

3.1 flip-flop 算法

我们假设特征矩阵 $\boldsymbol{X}$ 服从矩阵正态分布 $\boldsymbol{X} \sim \mathcal{MN}_{d_1,\ d_2}(\boldsymbol{M},\ \boldsymbol{\Sigma}_r,\ \boldsymbol{\Sigma}_c)$，其协方差矩阵形式为：

$$\mathrm{Var}\left[\mathrm{vec}(\boldsymbol{X})\right] = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r. \tag{3.9}$$

这种协方差形式（3.9）将（2.5）中的目标函数简化为

$$L(\boldsymbol{u}, \boldsymbol{v}, t) = (\boldsymbol{u}^T\boldsymbol{\Sigma}_r\boldsymbol{u})(\boldsymbol{v}^T\boldsymbol{\Sigma}_c\boldsymbol{v}) + \lambda\mathbb{E}\left\{1 - \tilde{Y}(\boldsymbol{u}^T(\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}))\boldsymbol{v} - t)\right\}_+. \tag{3.10}$$

在样本层面，我们需要估计特征矩阵的均值和协方差矩阵 $\boldsymbol{X}$，以用于目标函数 L（$\boldsymbol{u}$，$\boldsymbol{v}$，t）中的（3.10）。我们提出了一种用于估计的 flip-flop 算法。

令 $\{(\boldsymbol{X}_i,\ Y_i)\}_{i=1}^n$ 为 $(\boldsymbol{X},\ Y) \in \mathbb{R}^{d_1\times d_2} \times \mathbb{R}$ 的独立同分布样本。给定样本矩阵 $\{\boldsymbol{X}_i\}_{i=1}^n$，我们通过以下方式估计均值矩阵和协方差矩阵：

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i$$
$$\hat{\boldsymbol{\Sigma}}_r = \frac{1}{d_2 n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})\hat{\boldsymbol{\Sigma}}_c^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T$$

$$\hat{\boldsymbol{\Sigma}}_c = \frac{1}{d_1 n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \hat{\boldsymbol{\Sigma}}_r^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}). \tag{3.11}$$

与均值参数不同，协方差参数没有闭式解，因为两个协方差矩阵相互依赖。因此，我们基于（3.11）迭代计算它们的估计值，直至收敛；这种方法称为“乒乓”算法（Dutilleul，1999；Glanz 和 Carvalho，2018）。注意，当特征矩阵服从矩阵正态分布时，（3.11）中的均值和协方差矩阵估计值即为最大似然估计量（MLE）。关于矩阵正态分布的最大似然估计量性质的详细介绍，可参见Ro´s 等人（(2016）。

关于 flip-flop 算法的特性和统计保证已有大量研究。众所周知，当且仅当 $n \geq \max(d_1/d_2, d_2/d_1) + 1$ (Dutilleul, 1999) 时，flip-flop 算法会收敛到正定协方差矩阵。近期，Franks 等人(2021) 给出了近乎最优的样本复杂度，并在条件 $n \geq C\frac{d_1}{d_1}\max\{\log d_2, \log^2 d_1\}$ 下确立了 flip-flop 算法的统计保证，其中 $C > 0$ 且 $1 < d_1 \leq d_2$。此外，他们将所有结果推广到了高阶张量情形。我们利用这些结果，并使用 flip-flop 算法的输出估计矩阵中心子空间。

—

3.2 PSMM 算法

现在，我们基于观测样本介绍用于矩阵充分降维的 PSMM 算法。暂且假设中心空间 r_1, r_2) 的结构维数（ $S_{Y|\boldsymbol{X}}$ 已知，即 $\dim(S^r_{Y|\boldsymbol{X}}) = r_1$ 且 $\dim(S^c_{Y|\boldsymbol{X}}) = r_2$。结构维数未知的情形将在第 3.3 节讨论。估计程序总结如下。

– ^ ^ 步骤 1. 使用翻转算法计算样本均值 $\boldsymbol{X}$ 和协方差矩阵（$\boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c$）

算法（3.11）。

步骤 2. 令 q_h 为样本 $\{Y_1, \ldots, Y_n\}$ 关于 $h \in [H]$ 的 h/H 分位数。定义 $\tilde{Y}_i^h = \mathbb{1}\{Y_i > q_h\} - \mathbb{1}\{Y_i \leq q_h\}$ for each $h \in [H]$.

步骤 3. 对于每个 $h \in [H]$，寻找使下式最小的解（$\tilde{\boldsymbol{u}}^h$，$\tilde{\boldsymbol{v}}^h$，$\tilde{t}^h$）

$$(\boldsymbol{u}^T \hat{\boldsymbol{\Sigma}}_r \boldsymbol{u})(\boldsymbol{v}^T \hat{\boldsymbol{\Sigma}}_c \boldsymbol{v}) + \frac{\lambda}{n} \sum_{i=1}^{n} \left\{ 1 - \tilde{Y}_i^h (\boldsymbol{u}^T (\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \boldsymbol{v} - t) \right\}_+ . \tag{3.12}$$

ˆ 步骤 4. 计算 $\boldsymbol{U}_n$ 的 r_1 个主特征向量（ˆ$\boldsymbol{u}_1$、ˆ$\boldsymbol{u}_2$、.....、ˆ$\boldsymbol{u}_{r_1}$）以及 $\boldsymbol{V}_n$ 的 r_2 个主特征向量（ˆ$\boldsymbol{v}_1$、ˆ$\boldsymbol{v}_2$、.....、ˆ$\boldsymbol{v}_{r_2}$），其中定义

$$\hat{\boldsymbol{U}}_n = \sum_{h=1}^{H} \tilde{\boldsymbol{u}}^h (\tilde{\boldsymbol{u}}^h)^T \quad \text{and} \quad \hat{\boldsymbol{V}}_n = \sum_{h=1}^{H} \tilde{\boldsymbol{v}}^h (\tilde{\boldsymbol{v}}^h)^T. \tag{3.13}$$

步骤 5. 通过下式估计中心子空间 $S_{Y|\boldsymbol{X}}$

$$\hat{S}_{Y|\boldsymbol{X}} = \text{span}(\{\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_{r_1}\}) \times \text{span}(\{\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_{r_2}\}).$$

步骤 3 在样本层面根据（3.10）优化 PSMM目标函数。该目标函数（3.12）是双凸的，即对于固定的 $\boldsymbol{v} \in \mathbb{R}^{d_2}$，它关于 $\boldsymbol{u}$ 是凸的；对于固定的 $\boldsymbol{u} \in \mathbb{R}^{d_1}$，它关于 $\boldsymbol{v}$ 是凸的。因此，我们使用坐标下降算法最小化方程（ 3.12 ），该算法在固定另一组参数的情况下，为一组参数求解凸优化问题。根据定理 2，我们迭代更新参数 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 。

第4步是基于主成分分析对齐列降维子空间和行降维子空间的分量。

定理2。 *1.* 若 $\boldsymbol{u}^*$ 使得 (3.12) 在 $\mathbb{R}^{d_1}$ 中达到最小值，其中 $\boldsymbol{v} \in \mathbb{R}^{d_2}$固定，则

$$\boldsymbol{u}^* = \frac{1}{2}\sum_{i=1}^{n}(\alpha_i^* \tilde{Y}_i^h)\frac{\hat{\boldsymbol{\Sigma}}_r^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{v}}{\boldsymbol{v}^T\hat{\boldsymbol{\Sigma}}_c\boldsymbol{v}},$$

其中，$(\alpha_1^*, \ldots, \alpha_n^*)$ 是该二次规划问题的解：

$$\begin{aligned} &minimize \ -\sum_{i=1}^{n}\alpha_i + \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\tilde{Y}_i^h\tilde{Y}_j^h\frac{\left((\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{v}\right)^T\hat{\boldsymbol{\Sigma}}_r^{-1}\left((\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{v}\right)}{\boldsymbol{v}^T\hat{\boldsymbol{\Sigma}}_c\boldsymbol{v}}, \\ &subject\ to \ \sum_{i=1}^{n}\alpha_i\tilde{Y}_i^h = 0 \ and \ 0 \le \alpha_i \le \frac{\lambda}{n} \ for \ all \ i \in [n]. \end{aligned}$$

2. 若 $\boldsymbol{v}^*$ 在(3.12) 下 $\mathbb{R}^{d_2}$ （固定 $\boldsymbol{u} \in \mathbb{R}^{d_1}$）取得最小值，则

$$\boldsymbol{v}^* = \frac{1}{2}\sum_{i=1}^{n}(\beta_i^* \tilde{Y}_i^h)\frac{\hat{\boldsymbol{\Sigma}}_c^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T\boldsymbol{u}}{\boldsymbol{u}^T\hat{\boldsymbol{\Sigma}}_r\boldsymbol{u}},$$

其中，$(\beta_1^*, \ldots, \beta_n^*)$ 是该二次规划问题的解：

$$\begin{aligned} &minimize \ -\sum_{i=1}^{n}\beta_i + \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\tilde{Y}_i^h\tilde{Y}_j^h\frac{\left((\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T\boldsymbol{u}\right)^T\hat{\boldsymbol{\Sigma}}_c^{-1}\left((\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T\boldsymbol{u}\right)}{\boldsymbol{u}^T\hat{\boldsymbol{\Sigma}}_r\boldsymbol{u}}, \\ &subject\ to \ \sum_{i=1}^{n}\beta_i\tilde{Y}_i^h = 0 \ and \ 0 \le \beta_i \le \frac{\lambda}{n} \ for \ all \ i \in [n]. \end{aligned}$$

证明。 定义 $\tilde{\boldsymbol{u}} = \hat{\boldsymbol{\Sigma}}_r^{1/2}\boldsymbol{u}$, $\tilde{\boldsymbol{v}} = \hat{\boldsymbol{\Sigma}}_c^{1/2}\boldsymbol{v}$, 以及 $\boldsymbol{Z}_i = \hat{\boldsymbol{\Sigma}}_r^{-1/2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\hat{\boldsymbol{\Sigma}}_c^{-1/2}$。那么 (3.12) 等价于

$$\|\tilde{\boldsymbol{u}}\|^2\|\tilde{\boldsymbol{v}}\|^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\left\{1 - \tilde{Y}_i^h(\tilde{\boldsymbol{u}}^T\boldsymbol{Z}_i\tilde{\boldsymbol{v}} - t)\right\}_+. \tag{3.14}$$

假设 $\tilde{\boldsymbol{v}} \in \mathbb{R}^{d_2}$ 为固定值。此时，最小化 (3.14) 等价于

$$\begin{aligned}&\min_{\tilde{\boldsymbol{u}},\boldsymbol{\xi},t} \|\tilde{\boldsymbol{u}}\|^2\|\tilde{\boldsymbol{v}}\|^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\xi_i,\\ &\qquad \text{subject to } \tilde{Y}_i^h(\tilde{\boldsymbol{u}}^T\boldsymbol{Z}_i\tilde{\boldsymbol{v}} - t) \geq 1-\xi_i \text{ and } \xi_i \geq 0 \text{ for all } i \in [n].\end{aligned}$$

拉格朗日原始函数为

$$L_P = \|\tilde{\boldsymbol{u}}\|^2\|\tilde{\boldsymbol{v}}\|^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left[Y_i^h(\tilde{\boldsymbol{u}}^T\boldsymbol{Z}_i\tilde{\boldsymbol{v}} - t) - (1-\xi_i)\right] - \sum_{i=1}^{n}\mu_i\xi_i,$$

我们对 $\tilde{\boldsymbol{u}}$，t，和 ξ_i 求其最小值。通过检查一阶条件可得

$$\begin{aligned}\tilde{\boldsymbol{u}} &= \sum_{i=1}^{n}\alpha_i\tilde{Y}_i^h\frac{\boldsymbol{Z}_i\tilde{\boldsymbol{v}}}{2\|\tilde{\boldsymbol{v}}\|^2},\\ 0 &= \sum_{i=1}^{n}\alpha_i\tilde{Y}_i^h,\\ \alpha_i &= \frac{\lambda}{n} - \mu_i \text{ for all } i \in [n],\end{aligned} \tag{3.15}$$

满足正约束 α_i，μ_i，$\xi_i \geq 0$ 对所有 $i \in [n]$ 均成立。将一阶条件代入拉格朗日原始函数，可得对偶目标函数为

$$\begin{aligned}\text{minimize } &-\sum_{i=1}^{n}\alpha_i + \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\tilde{Y}_i^h\tilde{Y}_j^h\frac{(\boldsymbol{Z}_i\tilde{\boldsymbol{v}})^T(\boldsymbol{Z}_i\tilde{\boldsymbol{v}})}{\|\tilde{\boldsymbol{v}}\|^2},\\ &\text{subject to } \sum_{i=1}^{n}\alpha_i\tilde{Y}_i^h = 0 \text{ and } 0 \leq \alpha_i \leq \frac{\lambda}{n} \text{ for all } i \in [n].\end{aligned}$$

将（3.15）中的参数代回原式，使用 $\tilde{\boldsymbol{u}} = \hat{\boldsymbol{\Sigma}}_r^{-1/2}\boldsymbol{u}$，$\tilde{\boldsymbol{v}} = \hat{\boldsymbol{\Sigma}}_c^{1/2}\boldsymbol{v}$，以及$\boldsymbol{Z}_i = \hat{\boldsymbol{\Sigma}}_r^{-1/2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\hat{\boldsymbol{\Sigma}}_c^{-1/2}$，完成了关于 $\boldsymbol{u}^*$. 更新 $\boldsymbol{v}$ （固定 $\boldsymbol{u}$）的第一部分，具体步骤相同，故略去。

□

3.3 确定结构维数

在实践中，我们需要估计 $S^r_{Y|\boldsymbol{X}}$和 $S^c_{Y|\boldsymbol{X}}$的结构维数。我们提出使用修正的贝叶斯信息准则（BIC）来估计上一节步骤 4 中的未知结构维数（r_1, r_2）：

$$\begin{aligned}\mathrm{BIC}(r_1) &= \sum_{i=1}^{r_1} \lambda_i(\hat{\boldsymbol{U}}_n) - \lambda_1(\hat{\boldsymbol{U}}_n) n^{-1/2} r_1, \\ \mathrm{BIC}(r_2) &= \sum_{i=1}^{r_2} \lambda_i(\hat{\boldsymbol{V}}_n) - \lambda_1(\hat{\boldsymbol{V}}_n) n^{-1/2} r_2,\end{aligned}$$

其中 $\lambda_i(\boldsymbol{M})$ 是矩阵 i 的第 $\boldsymbol{M}$ 大特征值对应的特征向量。我们选择使贝叶斯信息准则最小的结构维数 ($\hat{r}_1, \hat{r}_2$)。类似的准则已见于 Zhu 等人(2006)；Wang 等人(2008)；Li 等人(2011)；ArtemiouandDong(2016)。当式（3.10）的海森矩阵在某个最优点处正定时，估计的结构维数具有一致性。详情请参见第 6 节。

4 扩展到高阶张量

我们将矩阵充分降维扩展到高阶张量情形。假设我们有阶数为-K （d_1，……， d_K）维的特征张量和响应变量（$\mathcal{X}_i$， Y_i） $\in \mathbb{R}^{d_1 \times \cdots \times d_K} \times \mathbb{R}$，对所有 $i \in [n]$. 均成立。我们的目标是找到 K 个矩阵$\boldsymbol{U}_k \in \mathbb{R}^{d_k \times r_k}$ （$r_k < d_k$），使得对于 $k = 1$，……， K，

$$Y \perp\!\!\!\perp \mathcal{X} | \mathcal{X} \times_1 \boldsymbol{U}_1 \times_2 \cdots \times_K \boldsymbol{U}_K. \tag{4.16}$$

这样，我们就保留了特征张量 $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ 中用于预测响应变量Y的全部信息，而且只需使用降维后的特征维度 $\mathcal{X} \times_1 \boldsymbol{U}_1 \times_2 \cdots \times_K \boldsymbol{U}_K \in \mathbb{R}^{r_1 \times \cdots \times r_K}$.。注意，方程（4.16）简化为矩阵充分降维问题

（1.2）中的矩阵情形（$K = 2$）。与矩阵情形类似，我们将中心模态-k子空间记作$\mathcal{S}^k_{Y|\mathcal{X}}$，定义如下：

$$\mathcal{S}^k_{Y|\mathcal{X}} = \bigcap_{\{\boldsymbol{U}_k : Y \perp\!\!\!\perp \mathcal{X}|\mathcal{X}\times_1 \boldsymbol{U}_1 \times_2 \cdots \times_K \boldsymbol{U}_K\}} \operatorname{span}(\boldsymbol{U}_k)$$

子空间 $S_{Y|\mathcal{X}} = \bigtimes_{k=1}^K S^k_{Y|\mathcal{X}}$称为高阶张量的中心子空间。

我们提出主支撑张量机（PSTM），将 PSMM 推广到高阶张量情形。我们考虑 PSTM 在总体层面上的以下目标函数。

$$\begin{aligned} L(\{\boldsymbol{u}_k\}_{k=1}^K, t) = \operatorname{Var}(\mathcal{X} \times_1 \boldsymbol{u}_1 \times_2 \cdots \times_K \boldsymbol{u}_K) \\ + \lambda \mathbb{E}\left\{1 - \tilde{Y}\left((\mathcal{X} - \mathbb{E}(\mathcal{X})) \times_1 \boldsymbol{u}_1 \times_2 \cdots \times_K \boldsymbol{u}_K - t\right)\right\}_+, \end{aligned} \tag{4.17}$$

其中，我们将随机变量定义为 $\tilde{Y} = \mathbb{1}\{Y \in A_1) - \mathbb{1}\{Y \in A_2\}$，对于支撑集 Y 中任意不相交子集 A_1 和 A_2。

采用与定理*1*中类似的证明思路，我们可以证明以下定理。定理 *3.* 假设 $\mathbb{E}($ $\mathcal{X}|\mathcal{X}\times_1 \boldsymbol{U}_1 \times_2 \cdots \times_K \boldsymbol{U}_K)$ 是 $\mathcal{X}\times_1 \boldsymbol{U}_1 \times_2 \cdots \times_K \boldsymbol{U}_K$ 的多线性函数，其中 $\{\boldsymbol{U}_k\}_{k=1}^K$ 是如 *(4.16)* 中定义的矩阵。若 $(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_K^*, t^*)$ 在所有满足 $\in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_K} \times \mathbb{R}$ 的 $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K, t)$ 中使目标函数 *(4.17)* 取得最小值，则 $(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_K^*) \in S_{Y|\mathcal{X}}$.

定理 3为高阶张量的中心子空间估计提供了指导。我们通过最小化一系列主支撑张量机的目标函数，并汇总所有极小化解，来估计中心子空间。由于估计程序与矩阵情形非常相似，我们仅在此强调主要差异。

在样本层面估计中，步骤 3 中主支撑张量机的目标函数

章节3 变为:

$$\prod_{k=1}^{K}(\boldsymbol{u}_k^T\hat{\boldsymbol{\Sigma}}_k\boldsymbol{u}_k)+\frac{\lambda}{n}\sum_{i=1}^{n}\left\{1-\tilde{Y}_i^h\left((\mathcal{X}_i-\bar{\mathcal{X}})\times_1\boldsymbol{u}_1\times_2\cdots\times_K\boldsymbol{u}_K-t\right)\right\}_+, \quad (4.18)$$

ˆ其中 $\boldsymbol{\Sigma}_k$ 是基于张量正态模型通过 flip-flop 算法得到的，其协方差具有如下结构：$\text{Var}\left(\text{vec}(\mathcal{X})\right)=\boldsymbol{\Sigma}_1\otimes\cdots\otimes\boldsymbol{\Sigma}_K$。这里略去 flip-flop 算法的细节，但需要指出的是，高阶张量情形下的算法及其一致性可见 Franks 等人(2021) 第 2 节。为最小化目标函数（4.18），我们采用支持张量机（STM）算法。具体而言，设 $\boldsymbol{u}_k'=\hat{\boldsymbol{\Sigma}}_k^{1/2}\boldsymbol{u}_k$和 $\mathcal{Z}_i=(\mathcal{X}_i-\bar{\mathcal{X}})\times_1\hat{\boldsymbol{\Sigma}}_1^{-1/2}\times_2\cdots\times_K\hat{\boldsymbol{\Sigma}}_K^{-1/2}$. 那么，我们将（4.18）重写为:

$$\prod_{k=1}^{K}\|\boldsymbol{u}_k'^T\|^2+\frac{\lambda}{n}\sum_{i=1}^{n}\left\{1-\tilde{Y}_i^h\left(\mathcal{Z}_i\times_1\boldsymbol{u}_1'\times_2\cdots\times_K\boldsymbol{u}_K'-t\right)\right\}_+,$$

这就是在Kotsia 和 Patras (2011)；Kotsia 等人 (2012) 中提出的带秩-1约束的正则支持张量机的目标函数。因此，我们可以应用标准STM算法求解该优化问题。最后，我们对 $(\tilde{\boldsymbol{u}}_1^h,\ldots,\tilde{\boldsymbol{u}}_K^h)$ 中的优化解进行汇总，得到 (4.18) 中的 $h\in[H]$，并按照第 $\mathcal{S}_{Y|\mathcal{X}}$3节第 4–5 步的方法估计中心子空间。

5 数值分析

本节中，我们分析合成数据集和真实世界数据集，以展示 PSMM 的性能。

5.1 合成数据

我们将 PSMM 的性能与现有的矩阵充分降维方法进行比较。比较分为两个方面。首先，我们比较基于矩阵的充分降维方法与一种传统的基于向量的方法的性能。其次，我们将本方法的性能与现有的矩阵预测变量充分降维方法进行比较。

- 主支持向量机（PSVM）（Li 等人，2011；Artemiou and Dong，2016）使用支持向量机进行充分降维。我们将特征矩阵向量化，并应用这种基于向量的充分降维方法。

- 折叠切片逆回归（folded-SIR）（Li 等人，2010；Dingand Cook，2015）基于 Li （1991）提出的切片逆回归方法。折叠-SIR 将基于向量的方法推广到矩阵情形，并涉及一阶逆矩。

- 折叠方向回归（folded-DR）（Li 等人，2010）基于（Li和Wang，2007）提出的方向回归方法。折叠-DR 将基于向量的方法推广到矩阵情形，并涉及二阶逆矩。

我们使用以下模型：

$$\begin{aligned}
\text{Model 1}: &\quad Y = \exp(\boldsymbol{X}_{11}) + \boldsymbol{X}_{12} + \epsilon, \\
\text{Model 2}: &\quad Y = \boldsymbol{X}_{11}/\{0.5 + (\boldsymbol{X}_{12} + 1)^2\} + \epsilon, \\
\text{Model 3}: &\quad Y = \boldsymbol{X}_{11}(\boldsymbol{X}_{12} + \boldsymbol{X}_{21} + 1) + \boldsymbol{X}_{11} + \epsilon.
\end{aligned}$$

其中，特征矩阵 $\boldsymbol{X} \in \mathbb{R}^{d\times d}$ 独立同分布地抽取自 $\mathcal{MN}_{d,\ d}(0_{d\times d},\ \boldsymbol{I}_d,\ \boldsymbol{I}_d)$，噪声 ϵ 独立同分布地抽取自 $N(0,\ 0。2_2)$。特征矩阵 $\boldsymbol{X}$ 的中心子空间在模型 1 和 2 中为 $\text{span}(\{e_1\}) \times \text{span}(\{e_1, e_2\})$，而在模型 3 中为 $\text{span}(\{e_1, e_2\}) \times \text{span}(\{e_1, e_2\})$。我们改变样本量 $n \in \{100, 200, \ldots, 500\}$ 和矩阵维数 $d \in \{5, 10\}$. 在我们的算法中，我们选择 q_h 为用于 $h \in [H]$的样本 $\{Y_1, \ldots, Y_n\}$的第 (h/H) 百分位数。我们设置超参数 $H = 10$ 和 $\lambda = 100$.

我们采用 Li 等人 (2005、2010提出的距离度量来评估各模型的表现。具体而言，设 $\mathcal{S}^r$ 和 $\mathcal{S}^c$ 分别为真实的中心行子空间和列子空间，而 $\mathcal{S}^r$ 和 $\mathcal{S}^c$ 为其估计值。则估计误差定义为

$$\text{dist}(\mathcal{S}^r \otimes \mathcal{S}^c, \hat{\mathcal{S}}^r \otimes \hat{\mathcal{S}}^c) = \left\| \boldsymbol{P}_{\mathcal{S}^r \otimes \mathcal{S}^c} - \boldsymbol{P}_{\hat{\mathcal{S}}^r \otimes \hat{\mathcal{S}}^c} \right\|_F,$$

其中，$\boldsymbol{P}_{\mathcal{S}}$ 是子空间 $\mathcal{S}$ 上的正交投影，$\|\cdot\|_F$ 是矩阵的 Frobenius 范数。所有汇总统计量均取 20 次重复试验的平均值。

图 1 展示了模型 1–3 中各种充分降维方法的估计误差，这些误差是在不同样本量和特征矩阵维数下评估的。值得注意的是，在所有情形下，PSMM 算法的表现始终优于其他方法。我们证实，基于矩阵的方法明显优于基于向量的方法，因为在所有情形下，所有基于矩阵的方法的表现都优于 PSVM。此外，PSMM 的表现也优于其他基于矩阵的方法。这种改进背后的直觉与 Li 等人(2011) 中的解释类似。由于切片逆回归和降维方法往往会降低数据点中心附近切片均值的权重，因此不适用于回归函数在数据点中心附近估计得更准确的情况，而这种情况在许多场景中都很常见 Kutner 等人 (2004)。相比之下，PSMM 使用用于分隔的超平面

数据点，因此不会对中心附近的数据产生降权效应。最后，我们发现，随着样本量增加，所有算法都呈现多项式衰减模式，这意味着估计量具有一致性。我们还发现，随着矩阵维数增加，算法的表现往往会下降。这并不令人意外，因为矩阵维数越大，待搜索的空间也越大。

5.2 在 MRN-114 人类脑连接数据上的应用

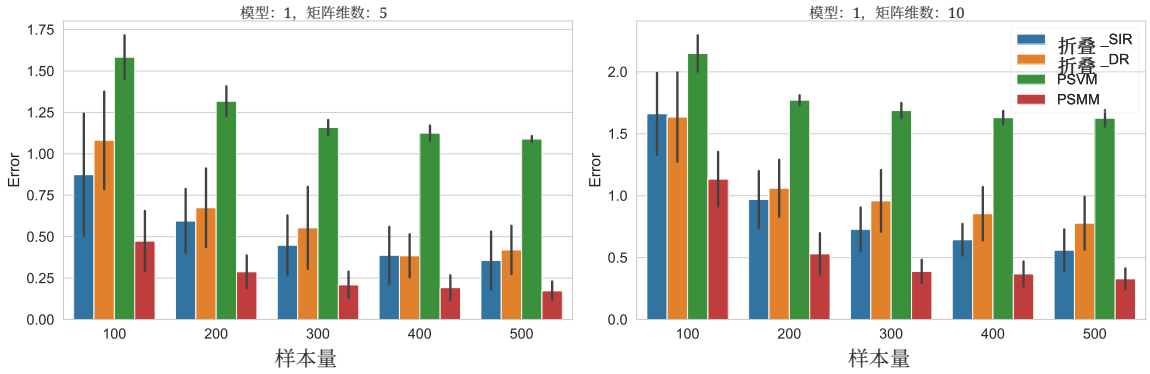
我们将 PSMM 应用于 MRN-114 数据集。该数据集包含 114 名受试者的 68 个脑节点的结构连接，以及通过全量表智商（FSIQ）分数测得的认知能力（Jung 和 Haier，2007；Wang 等，2017）。我们将连接数据转换为邻接矩阵$\boldsymbol{X}_i \in \mathbb{R}^{68\times 68}$，其中 $i \in [114]$的每个条目表示 68 个不同脑区之间是否存在纤维连接。每个邻接矩阵 $\boldsymbol{X}_i$ 对应的响应变量是 FSIQ 分数 Y_i，其范围为 86 至 144。我们使用输入 $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{114}$ 运行 PSMM 算法，并估计矩阵 $\boldsymbol{U} \in \mathbb{R}^{64\times r}$，使得 $Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{U}$。为便于可视化和解释，我们设定 $r = 2$。基于估计矩阵 $\hat{\boldsymbol{U}} = (\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \in \mathbb{R}^{64\times 2}$，我们为每个观测 $i \in [114]$计算降维特征变量 V_i^1、 V_i^2 和 V_i^3，定义如下：

$$V_i^1 = \hat{\boldsymbol{u}}_1^T \boldsymbol{X}_i \hat{\boldsymbol{u}}_1 = \langle \hat{\boldsymbol{u}}_1 \hat{\boldsymbol{u}}_1^T, \boldsymbol{X}_i \rangle,$$

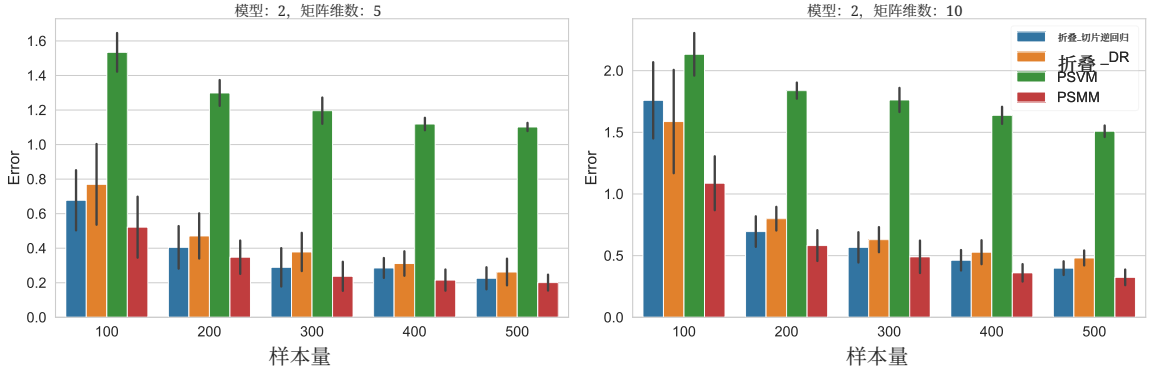
$$V_i^2 = \hat{\boldsymbol{u}}_2^T \boldsymbol{X}_i \hat{\boldsymbol{u}}_2 = \langle \hat{\boldsymbol{u}}_2 \hat{\boldsymbol{u}}_2^T, \boldsymbol{X}_i \rangle,$$

$$V_i^3 = \hat{\boldsymbol{u}}_1^T \boldsymbol{X}_i \hat{\boldsymbol{u}}_2 = \langle \hat{\boldsymbol{u}}_1 \hat{\boldsymbol{u}}_2^T, \boldsymbol{X}_i \rangle.$$

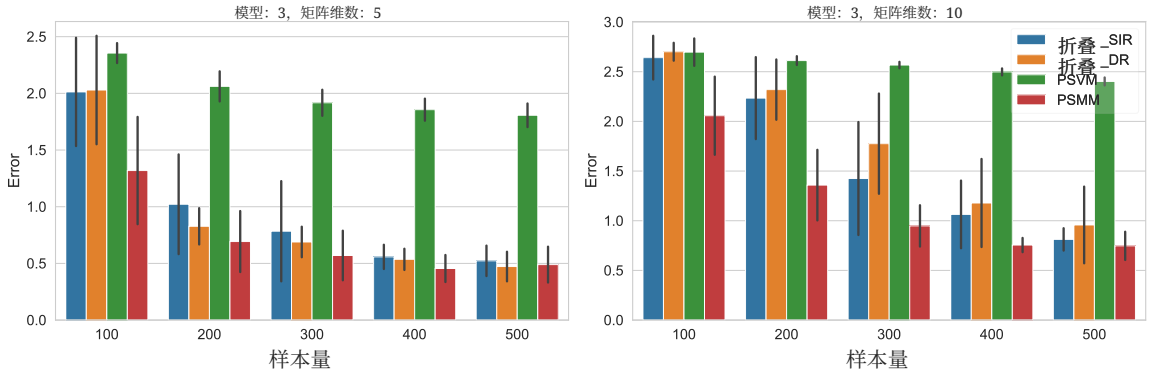
图2 按顺序展示了降维特征变量的载荷矩阵 V^1、 V^2和 V^3。我们发现，所有载荷矩阵都具有稀疏结构，这意味着只有部分脑网络能够显著解释全量表智商（FSIQ）得分。图3（a）绘制了



(a) 模型 1 的估计误差



(b) 模型 2 的估计误差



(c) 模型 3 的估计误差

图 1：四种方法在不同样本量和特征矩阵维度下的估计误差。

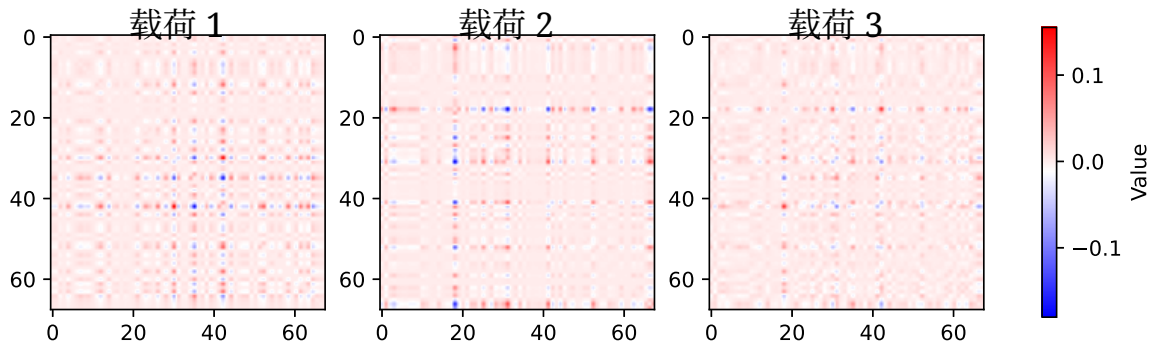
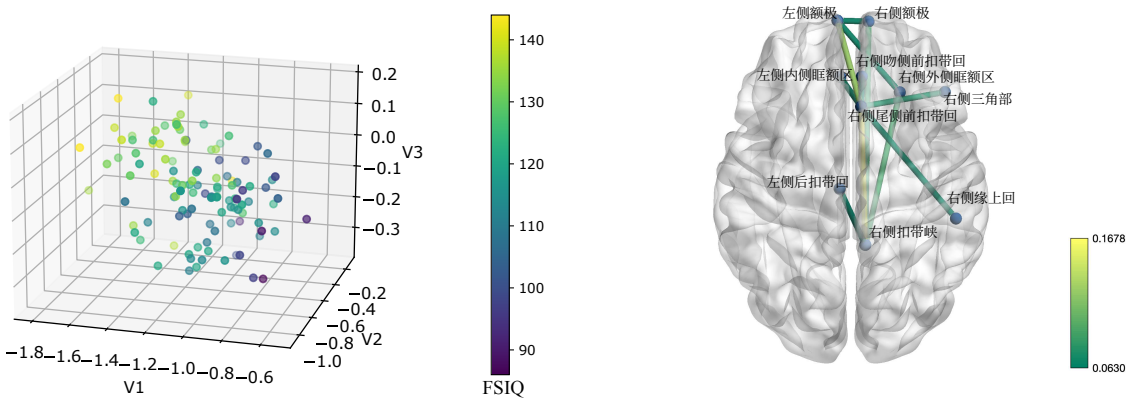


图 2：降维特征变量的载荷矩阵，依次为 V^1, V^1 和 V^3。

降维特征变量 V^1、V^2 和 V^3，以及个体的全量表智商分数。令人惊讶的是，这三个特征变量很好地捕捉了全量表智商的变化趋势。例如，V^1取较大的负值、V^2取较小的负值且 V^3 取正值的个体，往往全量表智商分数较高；而 V^1取较小的负值、V^2取较大的负值且 V^3 取负值的个体，则倾向于全量表智商分数较低。此外，我们检查了 V^1 的载荷矩阵中的元素（即 $\hat{\boldsymbol{u}}_1\hat{\boldsymbol{u}}_1^t$），并绘制了负值最大的前10条脑连接，如图 3(b)所示。有趣的是，一个名为右侧扣带峡部的脑节点与其他节点之间存在多条边。众所周知，该区域参与多种认知和情绪过程，并已被发现与认知功能的某些方面（包括智力）有关（Vogt 等，2006；Li 和 Tian，2014）。此外，我们观察到，除与右侧扣带峡部的连接外，这些连接大多为半球间连接。这一发现也与近期关于脑连接与智力相关性的研究一致（Wang 等，2017；Lee 和 Wang，2021）。



(a) (b)

图3: （a）PSMM估计的三个降维特征变量的散点图。点的颜色表示个体对应的全量表智商分数。（b）脑连接数据中与全量表智商相关性最高的10条边。

6 结论与讨论

我们提出了一种新的矩阵充分降维方法，称为主支撑矩阵机（PSMM）。与其他现有降维方法相比，PSMM 保留了预测变量的矩阵结构，并能更准确地估计中心子空间。数值分析表明，我们提出的 PSMM 有效且适用。

我们的工作还可以进行若干扩展。尽管我们观察到经验性证据表明，估计误差以多项式速率衰减并趋于零，但我们尚未证明估计量的统计收敛性。事实上，我们可以利用 Jiang 等人(2008)；Koo 等人 (2008) 关于支持向量机的渐近结果来构造一致性证明。假设目标函数的海森矩阵 $L(\boldsymbol{u},\ \boldsymbol{v},\ t)$ 在 (3.10) 的一个最优点 $(\boldsymbol{u}^*,\ \boldsymbol{v}^*,\ t^*)$ 处正定。那么，结合 Jiang 等人 (2008) 中定理2的类似证明思路，以及 $\boldsymbol{U}_n,\ \boldsymbol{V}_n)$ 的构造

(3.13) 得出

$$\hat{\boldsymbol{U}}_n - \boldsymbol{U} = \mathcal{O}_p(n^{-1/2}) \quad \text{and} \quad \hat{\boldsymbol{V}}_n - \boldsymbol{V} = \mathcal{O}_p(n^{-1/2}). \tag{6.19}$$

因此，根据 (6.19) 和 Bura 和 Pfeiffer(2008) 的结果，我们可以得到 PSMM 的一致性，即 $\hat{}\boldsymbol{U} - \boldsymbol{U} = \hat{}\boldsymbol{V} - \boldsymbol{V} = \mathcal{O}_p(n^{-1/2})$，其中 $\hat{}\boldsymbol{U}$ 和 $\hat{}\boldsymbol{V}$ 是 PSMM 的输出。然而，与向量情形不同，海森矩阵的正定性并无保证。具体而言，在某些技术条件下，我们证明 $L(\boldsymbol{u},\ \boldsymbol{v},\ t)$ 的海森矩阵具有如下显式形式

$$\boldsymbol{H} = \boldsymbol{H}_1 + \lambda \sum_{\tilde{y}=-1,1} (\boldsymbol{H}_2 + \boldsymbol{H}_3)\mathbb{P}\left[\tilde{Y} = \tilde{y}\right],$$

其中我们定义

$$\boldsymbol{H}_1 = 2\begin{pmatrix} (\boldsymbol{v}^T\boldsymbol{\Sigma}_c\boldsymbol{v})\boldsymbol{\Sigma}_r & 2\boldsymbol{\Sigma}_r\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{\Sigma}_c & 0_{d_1\times 1} \\ 2\boldsymbol{\Sigma}_c\boldsymbol{v}\boldsymbol{u}^T\boldsymbol{\Sigma}_r & (\boldsymbol{u}^T\boldsymbol{\Sigma}_r\boldsymbol{u})\boldsymbol{\Sigma}_c & 0_{d_2\times 1} \\ 0_{1\times d_1} & 0_{1\times d_2} & 0 \end{pmatrix},$$

$$\boldsymbol{H}_2 = \mathbb{E}\left[\begin{pmatrix} \boldsymbol{X}\boldsymbol{v} \\ \boldsymbol{X}^T\boldsymbol{u} \\ -1 \end{pmatrix}\begin{pmatrix} \boldsymbol{X}\boldsymbol{v} \\ \boldsymbol{X}^T\boldsymbol{u} \\ -1 \end{pmatrix}^T \middle| \boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v} = t + \tilde{y}\right] f_{\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v}|\tilde{Y}}(t+\tilde{y}),$$

$$\boldsymbol{H}_3 = -\mathbb{E}\left[\begin{pmatrix} 0 & \boldsymbol{X} & 0 \\ \boldsymbol{X}^T & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbb{1}\{\tilde{y}(\boldsymbol{u}^T\boldsymbol{X}\boldsymbol{v} - t) < 1\}\right].$$

这里 $f_{\cdot|\cdot}$ 表示条件概率密度函数。值得注意的是，由于支持向量机具有凸性，正定海森矩阵是其天然具备的性质。然而，由于 SMM 具有非凸性，检查其海森矩阵的正定性并非易事。寻找

使海森矩阵在最优点处为正定的显式条件，值得未来研究。

证明贝叶斯信息准则的一致性是另一个有趣的问题。在第3.3节中，我们提出了修正的贝叶斯信息准则，用于估计中心子空间的真实结构维数（r_1, r_2）。我们可以证明，在方程（2.8）的海森矩阵在最优点处正定这一假设下，贝叶斯信息准则具有一致性。下面简要概述证明。如上所述，正定的海森矩阵保证了根据方程（6.19）得到的中心子空间估计量的一致性。在这一一致性基础上，我们可以沿用 Li 等（2011）中定理8的证明思路。最后，我们将 Li 等（2011）中定理8的常数设为 $c_1(n) = n^{1/2} \log n$ 和 $c_2(k) = k$，从而完成贝叶斯信息准则一致性的证明，即 $\lim_{n\to\infty} \mathbb{P}(\hat{r}_1 = r_1)$ 和 $\lim_{n\to\infty} \mathbb{P}(\hat{r}_2 = r_2)$。

参考文献

Artemiou, A. 和 Y. Dong (2016)。通过主成分左 q支持向量机实现充分降维。《电子统计学杂志》*10*(1), 783–805。

Bura, E. 和 R. Pfeiffer (2008)。关于随机矩阵左奇异向量的分布及其应用。《统计与概率快报》*78*(15), 2275– 2280。

Chiaromonte, F. 和 R. D. Cook (2002)。回归中的充分降维与图形。《统计数学研究所年刊》*54*, 768–795。

Cook, R. D. 和 L. Ni（2005）。通过逆回归实现充分降维：一种最小差异方法。《美国统计学会会刊》*100*(470)，410–428。

Cook, R. D. 和 S. Weisberg（1991）。关于用于降维的切片逆回归的讨论。《美国统计学会会刊》*86*(414)，328–332。

Desikan, R. S., F. S´egonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. (2006). 一种自动标注系统，用于将 MRI扫描中的人类大脑皮层细分为基于脑回的感兴趣区域。*Neuroimage 31*(3), 968–980.

Ding, S. and R. D. Cook (2015). 张量切片逆回归。*Journal of Multi-variate Analysis 133*, 216–231.

Duan, N. and K.-C. Li (1991). 切片回归：一种无链接回归方法。统计年刊, 505–530.

Dutilleul, P. (1999). 矩阵正态分布的最大似然估计算法。统计计算与模拟期刊 *64*(2), 105–123.

Franks, C., R. Oliveira, A. Ramachandran, 和 M. Walter (2021). 利用测地凸性，为矩阵与张量正态模型实现近乎最优的样本复杂度。*arXiv* 预印本 *arXiv:2110.07583*。

Glanz, H. 和 L. Carvalho (2018). 矩阵正态分布的期望最大化算法及其在遥感中的应用。多变量分析期刊 *167*，31–48。

Jiang, B., X. Zhang, 和 T. Cai (2008). 估计支持向量机分类器预测误差的置信区间。机器学习研究期刊 *9*，521–540。

Jung, R. E. 和 R. J. Haier（2007）。

智力：趋同的神经影像学证据。《行为与脑科学 *30*》(2), 135–154。

Koo, J.-Y., Y. Lee, Y. Kim 和 C. Park（2008）。线性支持向量机的 Bahadur 表示。机器学习研究期刊 *9*，1343–1368。

Kotsia, I., W. Guo 和 I. Patras（2012）。用于视觉识别的高阶支持张量机。模式识别 *45*(12)，4192–4203。

Kotsia, I. 和 I. Patras（2011）。支持Tucker机。载于*CVPR 2011*，第 633–640 页。IEEE。

Kutner, M. H., C. J. Nachtsheim, J. Neter 和 W. Wasserman（2004）。应用线性回归模型，第 4 卷。McGraw-Hill/Irwin，纽约。

Lee, C. 和 M. Wang（2021）。超越符号：通过符号级数进行非参数张量补全。神经信息处理系统进展 *34*。

Li, B.、A. Artemiou 和 L. Li（2011）。用于线性和非线性充分降维的主支持向量机。统计年刊 *39*（6），3182–3210。

Li, B. 和 Y. Dong（2009）。非椭圆分布预测变量的降维。统计年刊 *37*（3），1272–1298。

Li, B.、M. K. Kim 和 N. Altman（2010）。关于矩阵或数组值统计对象的维度折叠。统计年刊 *38*（2），1094–1121。

Li, B. 和 S. Wang（2007）。关于用于降维的方向回归。《美国统计学会会刊》*102*（479），997–1008。

Li, B.、H. Zha 和 F. Chiaromonte（2005）。等高线回归：一种通用的降维方法。《统计年鉴》*33*，1580–1616。

Li, C. 和 L. Tian（2014）。儿童晚期和青春期顶叶-额叶网络静息态共同激活与智力之间的关联。《美国神经放射学杂志》*35*（6），1150–1156。

Li, K.-C. (1991). 用于降维的切片逆回归。《美国统计学会会刊》86(414), 316–327。

Luo, L., Y. Xie, Z. Zhang, and W.-J. Li (2015). 支持矩阵机。载于国际机器学习会议，第 938–947 页。

Pirsiavash, H., D. Ramanan, and C. C. Fowlkes (2009). 用于视觉识别的双线性分类器。载于《神经信息处理系统进展》，第 1482–1490 页。

Ro´s, B., F. Bijma, J. C. de Munck, and M. C. de Gunst (2016). 具有克罗内克积协方差结构的模型中最大似然估计量的存在性与唯一性。《多元分析杂志》143, 345–361。

Vogt, B. A., L. Vogt, and S. Laureys (2006). 人类后扣带区的细胞学特征及其功能相关神经回路。Neuroimage 29(2), 452–466。

Wang, J., X. Shen 和 Y. Liu（2008）。大间隔分类器的概率估计。《生物统计学》*95*(1), 149–167。

Wang, L., D. Durante, R. E. Jung 和 D. B. Dunson（2017）。贝叶斯网络–响应变量回归。《生物信息学》*33*(12), 1859–1866。

Yin, X., B. Li, and R. D. Cook（2008）。用于估计多指标回归中中心子空间的逐步方向提取。《多元分析杂志》*99*(8), 1733–1757.

Zhu, L., B. Miao, and H. Peng（2006）。关于高维协变量的切片逆回归。《美国统计学会会刊》*101*(474), 630–643.