



南開大學
Nankai University

《统计计算》 论文阅读报告

Bayesian Multidimensional Scaling and Choice of Dimension

姓 名：蒋贵豪

年 级：2021 级

学 院：统计与数据科学学院

专 业：应用统计学

完成日期：2021 年 12 月 7 日

摘 要

多维标度被广泛用于处理由对象对之间的相似性或不相似性度量组成的数据。我们解决了度量多维标度中的两个主要问题：一是对象的配置，二是对象维度的确定配置（在贝叶斯框架内）。在本文中，我们提出了用于对象配置的马尔可夫链蒙特卡罗算法，以及一个简单的贝叶斯标准，称为 MDSIC，用于选择维度。我们提供了使用我们的方法模拟结果以及真实数据的比较。我们的方法为对象配置提供比经典多维缩放和 ALSCAL 更好的结果。并且在我们的例子中，MDSIC 对于维度选择有较好的效果。

关键词：聚类；维度；差异性；马尔可夫链蒙特卡罗；度量尺度；模型选择

Abstract

Multidimensional scaling is widely used to handle data that consist of similarity or dissimilarity measures between pairs of objects. We deal with two major problems in metric multidimensional scaling—configuration of objects and determination of the dimension of object configuration—within a Bayesian framework. A Markov chain Monte Carlo algorithm is proposed for object configuration, along with a simple Bayesian criterion, called MDSIC, for choosing their dimension. Simulation results are presented, as are real data. Our method provides better results than does classical multidimensional scaling and ALSCAL for object configuration, and MDSIC seems to work well for dimension choice in the examples considered.

Key Words: Clustering; Dimensionality; Dissimilarity; Markov chain Monte Carlo; Metric scaling; Model selection

目 录

1 背景介绍	1
2 经典多维标度 CMDS 和 ALSCAL	1
2.1 经典多维标度 CMDS	1
2.2 ALSCAL	3
3 贝叶斯多维标度	3
3.1 模型和先验	3
3.2 MCMC	4
3.3 后验推断	5
4 维度选择	6
5 实验	8
5.1 算法流程梳理	8
5.2 数据介绍	9
5.3 最优维度选择	9
5.4 收敛分析	10
6 小组分工	12

1 背景介绍

多维标度 (MDS) 解决数据中的数据对之间相似性或不相似度量问题。它通过欧氏空间中的一个点表示一个数据, 其最早来源于心理学。MDS 的最主要应用是可视化, 即以更直观的形式表现出一组复杂的差异。另一个应用是了解度量这种差异的一个合适的维度。MDS 中一个重要的问题是数据值的估计, 我们称为对象属性值的估计。即给定数据值差异的情况下, 还原出我们得到这个差异的原始的数据, 这里的原始数据就是对象属性值。现在已有经典的 MDS 方法, 称为 CMDS, 可以在给定差异数据情况下还原对象配置。并且给出的差异数据的度量为欧式距离, 且指定我们正确的度量维度时, 它可以完全还原对象配置 (最多存在位置偏移)。

另一种常见的 MDS 方法称为 ALSCAL, 它最小化平方差异和平方距离方差之和。但是, 由于测量差异时存在误差, 且我们没有办法指定具体维度, 该方法也存在一定的局限性。

现在也有一些极大似然 MDS 方法来处理测量误差, 但是其依赖于渐进理论。如果我们需要的维度较高, 所需要的参数量就会很多, 并且高维情况容易陷入局部极小, 因此极大似然 MDS 方法在高维情况下并不适用。

MDS 中另一个重要的问题就是维度, 也就是度量属性的数量。也就是说, 给定一组差异数据, 如何确定原始数据的维度。现在还没有确定不同数据维度的确定性方法。

本文中, 主要在贝叶斯框架下处理上面两个 MDS 中的重要问题。我们使用欧氏距离, 并假设观察到的不同点存在高斯测量误差。在该假设下, 我们提出了一种简单的马尔可夫链蒙特卡罗 (MCMC) 算法, 通过该算法可以获得贝叶斯框架下的对象配置, 我们称我们的算法为 BMDS。之后, 在 BMDS 对一系列维度上的对象配置估计的基础上, 提出了一个简单的贝叶斯标准, 用于选择合适的维度。该标准称为 MDSIC, MDSIC 基于似然比, 用在一个维度与其他维度下的 BMDS 估计量上。在我们的方法中, 观察到的差异被建模为等于欧氏距离加上测量误差。从这个意义上说, 我们在这里所做的可以看作是对度量 MDS 的贝叶斯分析, 这里的贝叶斯分析得到了良好估计配置, 并提供了一种选择正确维度方法且提供了估计中的不确定性度量。

2 经典多维标度 CMDS 和 ALSCAL

2.1 经典多维标度 CMDS

若 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ 为在 p 维空间中未观测到的向量。设不相似性的度量 δ_{ij} , 定义为:

$$\delta_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

现在我们已知 δ_{ij} , 但是 \mathbf{x}_i 未知, CMDS 的目的就是从 δ_{ij} 中还原出 \mathbf{x}_i 。

但是在我们的实际应用中，我们只有 δ_{ij} 的观测值 d_{ij} ，其可能存在一定的测量误差。现在我们构建双中心矩阵 A ， A 中的元素 a_{ij} 定义为：

$$a_{ij} = -\frac{1}{2} (\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2) \quad (2)$$

其中， $\delta_{i.}^2 = \frac{1}{n} \sum_{j=1}^n d_{ij}^2$ ， $\delta_{.j}^2 = \frac{1}{n} \sum_{i=1}^n d_{ij}^2$ ， $\delta_{..}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$ 。

现在已经有定理给出：

$$a_{ij} = \sum_{k=1}^p x_{ik}x_{jk} \quad \text{i.e.} \quad A = \mathbf{X}\mathbf{X}^\top \quad (3)$$

其中， \mathbf{X} 是一个 $n \times p$ 维的矩阵，它在 p 维空间中代表对象的原始属性值。但是，实际应用中，往往我们只知道不相似性的度量 δ_{ij} ，而并不知道其对应的 \mathbf{X} ，甚至都不知道 \mathbf{X} 对应的维度 p 。因此我们需要做的就是从 δ_{ij} 中得到 A ，再运用谱分解定理还原 \mathbf{X} ，以及寻找 \mathbf{X} 的一个合适维度 p 。

但是我们从式2和式3中看到，我们在还原 \mathbf{X} 的过程中，用到的是观测数据 d_{ij} ，而不是真实数据 δ_{ij} 。如果我们的测量误差较大， d_{ij} 会远远偏离 δ_{ij} ，此时，CMDS 方法就不可用了。

我们知道，欧式距离在平移，旋转和绕原点的反射下是不变的。因此我们在上述意义下，可以将 \mathbf{X} 以原点为中心并旋转到主轴方向，从而方便我们的计算。

下面我们考虑维度 p 的选择，CMDS 中没有确定的方法来选取较好的维度 p ，即对差异有显著贡献的对象属性的数量。在 CMDS 中，评估维度最常用方法是看式2中矩阵 A 的特征值。其中第 k 个特征值是 \mathbf{X} 的第 k 个坐标对平方距离的贡献的度量。因此，只有对应于 \mathbf{X} 的最大的前 p 个特征值来表示对象差异。而要确定前 p 个特征值，可以绘制顺序特征值与维度的关系图，并寻找特征值序列趋于平稳的维度。如果每个 δ_{ij} 等于式1中给出的 p 维欧氏距离，那么在 $p+1$ 维处绘制的关系图应该是平稳的。

对于合适维度的度量，也可以通过计算一些指标来判断，比如常用的一种指标叫 STRESS：

$$\text{STRESS} = \sqrt{\frac{\sum_{i>j} (d_{ij} - \hat{\delta}_{ij})^2}{\sum_{i>j} d_{ij}^2}} \quad (4)$$

其中， $\hat{\delta}_{ij}$ 是 δ_{ij} 的一个估计量。在真实维度 p 处，STRESS 和维度的关系图也会趋于平稳。但是，无论是上述特征值方法还是指标方法，都依赖于绘图寻找平稳点。在我们真实使用中，如果数据和模型要求有些偏差，或者测量存在误差，又或者数据较多，累计偏差较大时，该方法并不适用。

2.2 ALSCAL

不同于 CMDS, ALSCAL 优化的是:

$$\text{S-STRESS} = \sum_{i>j} (\delta_{ij}^2 - d_{ij}^2)^2 \quad (5)$$

其与 STRESS 的不同是它为了计算方便, 使用了平方距离。但是, 这会导致 S-STRESS 强调较大差异而忽视较小差异, 这也是 ALSCAL 的一个缺点。

3 贝叶斯多维标度

3.1 模型和先验

在实际中, 我们在观测差异性时候会存在误差, 并且相异性度量一般都取正值。因此, 我们假设观测到的差异度量 d_{ij} 等于真实度量 δ_{ij} 加上一项高斯误差, 且限制差异度量为正。即给定 δ_{ij} , 假设观测到的差异度量 d_{ij} 遵循截断的正态分布:

$$d_{ij} \sim N(\delta_{ij}, \sigma^2) \mathbb{I}(d_{ij} > 0), \quad i \neq j, i, j = 1, \dots, n \quad (6)$$

其中, $\delta_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$, 且 x_{ik} 是我们的未知量。我们可以得到我们未知参数 \mathbf{X} 和 σ^2 的极大似然函数:

$$\begin{aligned} l(\mathbf{X}, \sigma^2) &= \prod_{i>j} \frac{(\sqrt{2\pi\sigma^2})^{-1} \exp\left(-\frac{(d_{ij} - \delta_{ij})^2}{2\sigma^2}\right)}{1 - \Phi\left(\frac{0 - \delta_{ij}}{\sigma}\right)} \propto \prod_{i>j} \frac{\frac{1}{\sigma} \exp\left(-\frac{(d_{ij} - \delta_{ij})^2}{2\sigma^2}\right)}{\Phi\left(\frac{\delta_{ij}}{\sigma}\right)} \\ &= (\sigma^2)^{-\frac{m}{2}} \exp\left(-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right) \end{aligned} \quad (7)$$

其中, $m = n(n-1)/2$ 为不相似性度量的个数。 $SSR = \sum_{i>j} (d_{ij} - \delta_{ij})^2$ 为残差平方和, $\Phi(\cdot)$ 为标准正态分布的累计密度函数。

在贝叶斯估计中, 我们需要有被估参数 \mathbf{X} 和 σ^2 的先验。对于 \mathbf{x}_i 的先验, 我们将其设为均值为 $\mathbf{0}$, 协方差矩阵为对角阵 Λ 的多元正态分布, 即: $\mathbf{x}_i \sim N(\mathbf{0}, \Lambda)$ 。对于方差 σ^2 的先验, 我们知道, 正态分布的方差的共轭先验分布是倒 Gamma 分布。于是, 我们设 σ^2 的先验: $\sigma^2 \sim IG(a, b)$ 。在给定维度 p 情况下, 对于超先验 $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$, 同样由正态分布的方差的共轭先验分布是倒 Gamma 分布, 我们设 λ_j 的先验为 $\lambda_j \sim IG(\alpha, \beta_j)$, 且对于 $j = 1, 2, \dots, p$ 是独立的。我们假设各先验 \mathbf{X} , Λ 和 σ^2 是独立的, 即: $\pi(\mathbf{X}, \sigma^2, \Lambda) =$

$\pi(\mathbf{X})\pi(\sigma^2)\pi(\Lambda)$ 。其中 $\pi(\mathbf{X})$, $\pi(\sigma^2)$, $\pi(\Lambda)$ 是我们刚刚给定的先验分布。

当我们只有较少的先验信息时,我们可以用我们在第二章节中的 CMDS 或 ALSCAL 给先验参数一个取值。例如,对于 σ^2 ,我们可以给定一个较小的参数 a ,然后计算 CMDS 或 ALSCAL 得到的残差均值 SSR/m ,那么我们的先验 b 就可以取使得倒 Gamma 分布均值 $b/(a-1)$ 等于残差均值 SSR/m 的 b 。同样对于超先验 λ_j ,我们也可以先给定一个较小的 α ,选取 β_j 使得 λ_j 的均值 $\beta_j/(\alpha-1)$ 和样本协方差阵 \mathbf{S}_x 的第 j 个对角元素相等,其中 $\mathbf{S}_x = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top \mathbf{x}_i$ 。这里,因为欧式距离的旋转、平移、中心对称不变性。

为方便我们的计算,我们取 $\sum_{i=1}^n \mathbf{x}_i = 0$,且其协方差矩阵为对角阵。当然,取这种先验对数据有一定的依赖性,是我们在实际使用中迫不得已的方法。

3.2 MCMC

为计算 $(\mathbf{X}, \sigma^2, \Lambda)$ 的后验分布,我们先写出先验 $\pi(\mathbf{X})$, $\pi(\sigma^2)$, $\pi(\Lambda)$ 的密度函数:

$$\pi(\mathbf{X}) = \prod_{i=1}^n \prod_{j=1}^p \left(\frac{1}{\sqrt{2\pi}\lambda_j} \right)^n \exp \left(-\frac{\mathbf{x}_i^\top \Lambda^{-1} \mathbf{x}_i}{2} \right) \propto \prod_{j=1}^p \lambda_j^{-\frac{n}{2}} \times \exp \left(-\frac{\sum_{i=1}^n \mathbf{x}_i^\top \Lambda^{-1} \mathbf{x}_i}{2} \right) \quad (8)$$

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)(\sigma^2)^{a+1}} \exp \left(-\frac{b}{\sigma^2} \right) \propto (\sigma^2)^{-(a+1)} \exp \left(-\frac{b}{\sigma^2} \right) \quad (9)$$

$$\pi(\Lambda) = \prod_{j=1}^p \frac{\beta_j^\alpha}{\Gamma(\alpha)(\lambda_j)^{\alpha+1}} \exp \left(-\frac{\beta_j}{\lambda_j} \right) \propto \prod_{j=1}^p (\lambda_j)^{-(\alpha+1)} \times \exp \left(-\sum_{j=1}^p \frac{\beta_j}{\lambda_j} \right) \quad (10)$$

结合式7、式8、式9和式10,我们便可以得到给定观测到的差异矩阵 \mathbf{D} 的 $(\mathbf{X}, \sigma^2, \Lambda)$ 的后验分布:

$$\begin{aligned} \pi(\mathbf{X}, \sigma^2, \Lambda | \mathbf{D}) &\propto (\sigma^2)^{-(m/2+a+1)} \prod_{j=1}^p \lambda_j^{-(n/2+\alpha+1)} \\ &\times \exp \left[-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi \left(\frac{\delta_{ij}}{\sigma} \right) - \frac{1}{2} \sum_{i=1}^n \mathbf{x}_i' \Lambda^{-1} \mathbf{x}_i - \frac{b}{\sigma^2} - \sum_{j=1}^p \frac{\beta_j}{\lambda_j} \right] \end{aligned} \quad (11)$$

这里,原论文中缺少一项: $\prod_{j=1}^p \lambda_j^{-(\alpha+1)}$

因为式11的形式过于复杂,且大多数情况下是高维的,因此我们采用 MCMC 的方法对式11进行模拟。我们通过交替更新对象配置 \mathbf{X} , 误差的方差 σ^2 和超先验 Λ , MCMC 得到的稳态分布就是我们的最终结果。

首先我们进行参数的初始化,对于 \mathbf{X} ,我们可以从 CMDS 或 ALSCAL 中得到一个

\mathbf{x}_i 作为我们的初始化 $\mathbf{x}_i^{(0)}$ 。因为欧式距离的平移旋转和中心对称不变性，为了我们的计算方便，我们将 $\mathbf{x}_i^{(0)}$ 中心化并进行谱分解，保证其所有元素和为 0，且有对角协方差矩阵。有了 $\mathbf{x}_i^{(0)}$ 后，我们便可以计算残差 $SSR^{(0)}$ 和 σ^2 的初始化： $\sigma^{2(0)} = SSR^{(0)}/m$ 。对于 Λ 的初始化，刚刚我们说明了 $\mathbf{X}^{(0)}$ 具有对角协方差矩阵，那么这个对角协方差矩阵的对角元素就可以当作 $\Lambda^{(0)}$ 。

下面说明 MCMC 中 \mathbf{X} ， Λ 和 σ^2 的更新方法。首先是 Λ 的更新，之前我们提到， λ_j 的共轭先验分布是倒 *Gamma* 分布，且 λ_j 的先验为 $\lambda_j \sim IG(\alpha, \beta_j)$ ，那么 λ_j 的后验分布为 $IG(\alpha + n/2, \beta_j + s_j/2)$ ，其中 $s_j = \sum_{i=1}^n (x_{ij} - 0)^2$ 。我们直接在后验分布中抽取，得到 Λ 的更新。

然后是 \mathbf{X} 的更新，我们使用 MCMC 中的 Metropolis-Hastings 算法来进行更新，我们的提议分布采用的是正态分布。正态分布的均值是上一步中 \mathbf{x}_i 的值，为了得到正态分布的方差，我们分析式11中自然对数指数部分中含有 \mathbf{x}_i 的项： SSR 和 $\mathbf{x}_i' \Lambda^{-1} \mathbf{x}_i$ 中是关于 \mathbf{x}_i 的二次项， $\log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)$ 是关于 \mathbf{x}_i 的一次项。为了方便，我们舍弃一次项。而在二次项中，对于每一个 i ， SSR 中含有 $n-1$ 个二次项，而 $\mathbf{x}_i' \Lambda^{-1} \mathbf{x}_i$ 中只有一个二次项。因此，我们最终认为 SSR 占支配地位。因此，自然对数的指数部分可以近似认为一个常数 $1/C$ 乘上 $(n-1)/(2\sigma^2)$ 。因此，我们最后的提议分布选择： $N(\mathbf{x}_i, C \frac{\sigma^2}{n-1})$ 。

最后是 σ^2 的更新，我们也使用 Metropolis-Hastings 算法来进行更新。之前我们提到， σ^2 的共轭先验分布是倒 *Gamma* 分布，且 σ^2 的先验为 $\sigma^2 \sim IG(a, b)$ ，那么 σ^2 的后验分布为 $IG(a + m/2, b + SSR/2)$ 。由于当 m 很大时，倒 *Gamma* 分布会近似于一个正态分布。于是，我们的提议正态分布的均值为上一步中 σ^2 的值，提议正态分布方差与 $IG(a + m/2, b + SSR/2)$ 的方差 $\frac{(b + SSR/2)^2}{(a + m/2 - 1)^2(a + m/2 - 2)}$ 成正比。

3.3 后验推断

我们将马氏链跑足够长的时间，得到 \mathbf{X} ， Λ 和 σ^2 的后验分布的估计。但是，正如之前所的，欧氏距离在平移，旋转和原点对称具有不变性，我们得到的 \mathbf{X} 只是相对位置正确的 \mathbf{X} ，而不是绝对位置正确的 \mathbf{X} 。因此，在我们跑马氏链的过程中，我们需要看的是 δ_{ij} 的收敛，而不是 \mathbf{X} 的收敛。很多情况下，即使 \mathbf{X} 不稳定， δ_{ij} 可以是稳定的。因此，我们将 \mathbf{X} 的后验近似估计作为 \mathbf{X} 的贝叶斯估计（即 BMDS 得到的解）。在适当的条件下，我们可以对得到的 \mathbf{X} 进行平移，旋转和中心对称变换，使得我们的结果有意义。

为了得到 \mathbf{X} 的后验，我们可以计算每个 MCMC 样本的后验密度。然而，这可能非常耗时，因为后验非常的复杂。然而，在我们上一节的讨论中，最大似然估计式7中， SSR 占主导地位，因此 \mathbf{X} 的后验可以靠最小化残差平方和 SSR 的 \mathbf{X} 来近似。

因为 \mathbf{X} 的中心和方向都可以是任意的。所以为了我们计算的方便， \mathbf{X} 的 MCMC 样本需要进行一定的变换处理。处理方式为： $\mathbf{x}_i = \mathbf{D}_x^T (\mathbf{x}_i - \bar{\mathbf{x}})$ 。其中， $\bar{\mathbf{x}}$ 为原先 \mathbf{x}_i 的均

值, \mathbf{D}_x 为样本协方差矩阵 $\mathbf{S}_x = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^\top (\mathbf{x}_i - \bar{\mathbf{x}})$ 的特征向量组成的矩阵。

尽管 \mathbf{X} 不稳定, 但是 δ_{ij} 经过 MCMC 后是稳定的。因此, 我们可以通过 δ_{ij} 的函数来对一些不确定性进行度量。

4 维度选择

在我们之前的讨论中, BMDS 给的是维度 p 已知的情况下对象配置的结果。但是正常情况下, 我们的维度 p 是未知的, BMDS 只是人工选择了一个维度, 下面我们使用贝叶斯估计来给处一个最合适的维度 p 。

如果我们把维度 p 也看作变量, 那么我们的后验分布可以写为:

$$\begin{aligned}
 & \pi(\mathbf{X}, \sigma^2, \Lambda, p | \mathbf{D}) \\
 & \propto l(\mathbf{D} | \mathbf{X}, \sigma^2, \Lambda, p) \pi(\mathbf{X} | \Lambda, p) \pi(\sigma^2) \pi(\Lambda, p) \\
 & = (2\pi)^{-m/2} \sigma^{-m} \exp \left[-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi(\delta_{ij}/\sigma) \right] \\
 & \times (2\pi)^{-np/2} \prod_{j=1}^p \lambda_j^{-n/2} \exp \left[-\sum_{j=1}^p \frac{1}{2\lambda_j} s_j \right] \\
 & \times \Gamma(a)^{-1} b^a (\sigma^2)^{-(a+1)} \exp[-b/\sigma^2] \\
 & \times \Gamma(\alpha)^{-p} \prod_{j=1}^p \beta_j^\alpha \lambda_j^{-(\alpha+1)} \exp[-\beta_j/\lambda_j] \\
 & = A(p) \cdot h(\sigma^2, \mathbf{X}) \cdot g(\Lambda, \mathbf{X}, p)
 \end{aligned} \tag{12}$$

论文原文中, 式12第二行不正确, 我们进行了修正。并且式12最后一行中:

$$A(p) = (2\pi)^{-(m+np)/2} \Gamma(a)^{-1} b^a \Gamma(\alpha)^{-p} \prod_{j=1}^p \beta_j^\alpha \tag{13}$$

$$h(\sigma^2, \mathbf{X}) = (\sigma^2)^{-(m/2+a+1)} \times \exp \left[-(SSR/2 + b)/\sigma^2 - \sum_{i>j} \log \Phi(\delta_{ij}/\sigma) \right] \tag{14}$$

$$g(\Lambda, \mathbf{X}, p) = \prod_{j=1}^p \lambda_j^{-(n/2+\alpha+1)} \exp[-(s_j/2 + \beta_j)/\lambda_j] \tag{15}$$

我们采用贝叶斯方法来选择维度, 即选取合适的维度 p 使后验分布 $\pi(\mathbf{X}, p | \mathbf{D})$ 达到最大。对于式14, 作者在附录中给出了:

$$\int h(\sigma^2, \mathbf{X}) d\sigma^2 \approx (2\pi)^{1/2} (m/2)^{-1/2} \times (SSR/m)^{-m/2+1} \exp[-m/2] \tag{16}$$

对于式15, 我们对 λ_j 求积分:

$$\begin{aligned}
 \int g(\Lambda, \mathbf{X}, p) d\Lambda &= \int \prod_{j=1}^p \lambda_j^{-(n/2+\alpha+1)} \exp[-(s_j/2 + \beta_j)/\lambda_j] d\lambda_j \\
 &= \prod_{j=1}^p \int -\lambda_j^{-(n/2+\alpha-1)} \exp[-(s_j/2 + \beta_j)/\lambda_j] d\lambda_j^{-1} \\
 &= \prod_{j=1}^p \int -\left(\frac{s_j/2 + \beta_j}{\lambda_j}\right)^{n/2+\alpha-1} \exp[-(s_j/2 + \beta_j)/\lambda_j] \left(\frac{1}{s_j/2 + \beta_j}\right)^{n/2+\alpha} d\frac{s_j/2 + \beta_j}{\lambda_j} \\
 &= \Gamma^p(n/2 + \alpha)(-1)^p \prod_{j=1}^p (s_j/2 + \beta_j)^{-(n/2+\alpha)}
 \end{aligned} \tag{17}$$

在论文原文中, 作者漏了一项 $(-1)^p$ 。

于是, 结合式12、式16和式17我们可以写出后验分布 $\pi(\mathbf{X}, p | \mathbf{D})$ 为:

$$\begin{aligned}
 \pi(\mathbf{X}, p | \mathbf{D}) &= c \cdot A(p) \cdot \int h(\sigma^2, \mathbf{X}) d\sigma^2 \cdot \int g(\Lambda, \mathbf{X}, p) d\Lambda \\
 &= c \cdot A^*(p) \cdot (SSR/m)^{-m/2+1} \times \prod_{j=1}^p (s_j/2 + \beta_j)^{-(n/2+\alpha)}
 \end{aligned} \tag{18}$$

其中, c 为和 X 和 p 独立的常数, 且:

$$A^*(p) = A(p) \cdot (2\pi)^{1/2} (m/2)^{-1/2} \exp[-m/2] (-1)^p \Gamma^p(n/2 + \alpha) \tag{19}$$

下面我们来选取最合适的 p , 令 $\mathbf{X}^{*(p+1)} = (\mathbf{X}^{(p)} : \mathbf{0})$, 即 $\mathbf{X}^{*(p+1)}$ 前 p 维为 $\mathbf{X}^{(p)}$, 最后一维为 0。我们知道, $\mathbf{X}^{*(p+1)}$ 和 $\mathbf{X}^{(p)}$ 在 $p+1$ 维中的欧氏距离是相等的。因此, 如果 p 是最好的维度, 那么在 $p+1$ 维中 \mathbf{X} 的最好选择就是 $\mathbf{X}^{*(p+1)}$ 。因此, 我们只需要计算 $\mathbf{X}^{*(p+1)}$ 和 $\mathbf{X}^{(p+1)}$ 后验的比值并和 1 作比较即可选出最合适的维度 p 。具体地, 我们对式18中计算两者的比值:

$$\begin{aligned}
 R_p &\equiv \frac{\pi(\mathbf{X}^{(p+1)}, p+1 | \mathbf{D})}{\pi(\mathbf{X}^{*(p+1)}, p+1 | \mathbf{D})} \\
 &= \left(\frac{SSR_{p+1}}{SSR_{p+1}^*}\right)^{-m/2+1} \left(\prod_{j=1}^{p+1} \frac{s_j/2 + \beta_j}{s_j^*/2 + \beta_j}\right)^{-(n/2+\alpha)} \\
 &= \left(\frac{SSR_{p+1}}{SSR_p}\right)^{-m/2+1} \left(\prod_{j=1}^p \frac{s_j^{(p+1)}/2 + \beta_j}{s_j^{(p)}/2 + \beta_j}\right)^{-(n/2+\alpha)} \\
 &\quad \times \left(\frac{s_{p+1}^{(p+1)}/2 + \beta_{p+1}}{\beta_{p+1}}\right)^{-(n/2+\alpha)}
 \end{aligned} \tag{20}$$

可以看出 R_p 的选取和 α, β_j 有关。

在式20中, 为了简化我们的计算, 我们将 α 取为 0.5, β_j 取为 $\beta_j = \frac{1}{2}s_j^{(p+1)}/n$, 那么式20变为:

$$R_p = \left(\frac{SSR_{p+1}}{SSR_p} \right)^{-m/2+1} \left(\prod_{j=1}^p \frac{r_j^{(p+1)}(n+1)}{\left(n + r_j^{(p+1)} \right)} \right)^{-(n+1)/2} \times (n+1)^{-(n+1)/2} \quad (21)$$

其中: $r_j^{(p+1)} = s_j^{(p+1)}/s_j^{(p)}$ 。我们对式21取对数, 并乘以 -2 :

$$\begin{aligned} LR_p &\equiv -2 \log R_p \\ &= (m-2) \log (SSR_{p+1}/SSR_p) \\ &\quad + \left\{ (n+1) \sum_{j=1}^p \log \left[\frac{r_j^{(p+1)}(n+1)}{\left(n + r_j^{(p+1)} \right)} \right] + (n+1) \log(n+1) \right\} \end{aligned} \quad (22)$$

当 LR_p 为正时, 说明 p 维度比 $p+1$ 好。反之, 若 LR_p 为负时, 说明 $p+1$ 维度比 p 好。因此我们设计一个指标, 称作 MDSIC:

$$\begin{aligned} \text{MDSIC}_1 &= (m-2) \log SSR_1 \\ \text{MDSIC}_p &= \text{MDSIC}_1 + \sum_{j=1}^{p-1} LR_j \end{aligned} \quad (23)$$

选取最小的 MDSIC_p , 此时的 p 就是我们维度的最优选择。

5 实验

5.1 算法流程梳理

对于整个算法, 我们首先利用已经给出的计算方法去构建相关函数: 定义出 δ 距离, SSR 残差平方和, 以及后验密度函数。由于 \mathbf{X} 是我们所要计算的, 其未知但我们知道它的后验分布。对于 \mathbf{X} 的先验参数, 我们可以按照前面的内容中给出, 然后我们利用后验分布结合 Metropolis-Hastings 算法进行迭代, 不断调整其中的参数使其逐渐收敛, 最后马氏链会收敛到我们的后验分布, 也就是我们 \mathbf{X} 和先验参数的分布函数。

算法的简要过程如下:

首先根据参数服从的分布预先确定一个参数初始值, 我们有观察到的距离矩阵, 并用之前 CMDS 得到的结果作为 \mathbf{X} 的初始化 (没有的话随机初始化)。通过 \mathbf{X} 我们计算残差平方和 SSR_0 作为 SSR 的初始化, 并利用 SSR_0/m 来表示方差的初始化。对于 λ 的初始化, 我们从其服从的倒伽马函数中抽取随机数作为初始值。选取完初始值之后我们利用 Metropolis-Hastings 算法来进行迭代随机模拟: 首先迭代产生 λ , 调整参数后生

成 \mathbf{X} , 利用 \mathbf{X} 计算 SSR/m 来迭代产生方差, 并形成完整循环。最终, 当马氏链多次迭代后趋于收敛, 去掉前面的燃烧期后, 后面的数据即为我们想要的稳定的符合后验分布的数据。对于其中的先验初始值, 我们选择 $a = 3, \alpha = 0.5$, 其余初值在上述内容已给定, 或可以由 a 和 α 计算出。由于马氏链的性质, 对于其中的一些先验参数, 例如 a, b, α, β_j 等先验参数的选取, 即其初始化的值不影响最终结果的收敛性及收敛值。

5.2 数据介绍

贵州茅台酒股份有限公司 (600519) 主要业务是茅台酒及系列酒的生产与销售。主导产品“贵州茅台酒”是世界三大蒸馏名酒之一, 也是集国家地理标志产品、有机食品和国家非物质文化遗产于一身的白酒品牌。贵州茅台股价在 A 股市场“一骑绝尘”, 因而其在 A 股时长有极高的关注度。

我们选取了贵州茅台 (600519) 39 个交易日的股票振幅、总手、金额、换手率和成交次数 5 项指标数据。我们直接从同花顺电脑端软件中直接导出了这些所需要的数据。在我们的数据预处理时, 因为我们所需要的数据从专业的证券软件中导出, 我们的数据并无异常, 并不需要进行数据清洗。数据详见“600519.csv”。

事实上, 在 39 个交易日中振幅、总手、金额、换手率和成交次数 5 项指标数就是我们的 \mathbf{X} 。这里, 我们的 \mathbf{X} 已知, 自然可以得出我们的天与天之间的指标相似性 δ_{ij} 。正常情况下, \mathbf{X} 是未知的, δ_{ij} 已知, 我们需要用 CMDS 或 ALSCAL 还原出 \mathbf{X} 。请注意, 上述还原 \mathbf{X} 过程中, 我们认为还原的维度 p 为原始维度。如果我们要进行最优维度选择, 那么我们还原得出的 \mathbf{X} 和我们的原始数据是不匹配的。

5.3 最优维度选择

利用文中给出的方法, 定义相关的 MDSIC 函数。我们得到不同维度下 MDSIC 的值, 并利用选择 MDSIC 取最小值的 p 为最优的维度。表2是 MDSIC 的计算结果, 我们迭代 1000 次, 确保马氏链已经收敛。

表 1: 各维度 MDSIC 结果

维度	1	2	3	4	5
MDSIC	34115.78	34240.11	34388.44	34520.18	34572.53

可以发现, MDSIC 的值随着 p 的变大不断变大, 在 $p = 1$ 处有最小, 之后越来越大, 于是维度为 1 时是最优的维度。

类比 PCA 选取主成分的过程, 我们对于这个得到的结果简要分析。由于我们是采用股票的交易数据来进行 BMDS, 而股票数据由于属于时间序列数据, 和时间存在着巨大的相关性, 因此数据降维后最优维度选择 1 维也和我们的常识保持一致。

5.4 收敛分析

当确定最优维度为 1 后，我们迭代 1000 次，对这个马氏链进行更为详尽的分析和介绍。首先我们绘制出 \mathbf{X} 的马氏链，如图1所示：

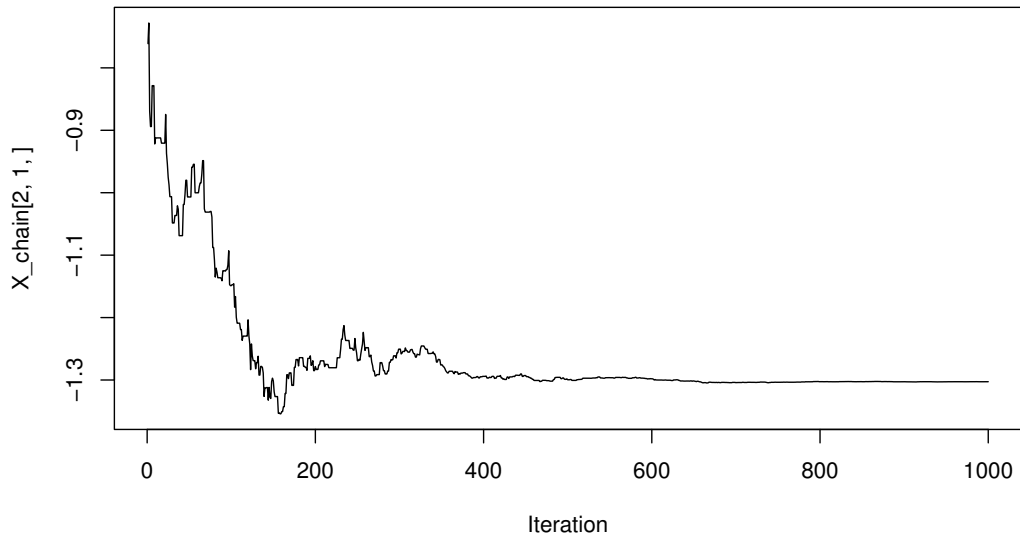


图 1: \mathbf{X} 的马氏链

通过图1可以发现，我们得到的关于 \mathbf{X} 的马氏链随着迭代次数的增加，最终在经过 350 次左右的迭代后已经很好的收敛到了一个小的区间范围里面。

然后我们绘制出后验分布参数方差 σ^2 的马氏链，如图2所示：

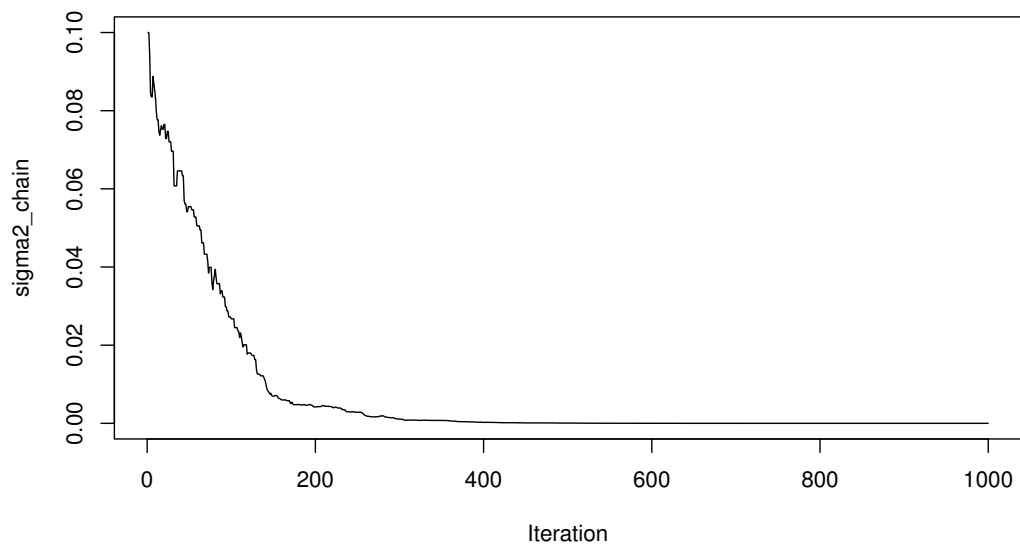


图 2: σ^2 的马氏链

由于 σ^2 是代表 \mathbf{X} 的方差，我们的目的就是使其尽可能的小，通过图像可以发现，我们得到的方差随着迭代次数的增多逐渐变小，很好的达成了我们尽可能降低方差目的。也就是说我们得到的数据越来越稳定到收敛值附近，波动也越来越小。

最后，我们绘制出超先验 λ 的马氏链，如图3所示：

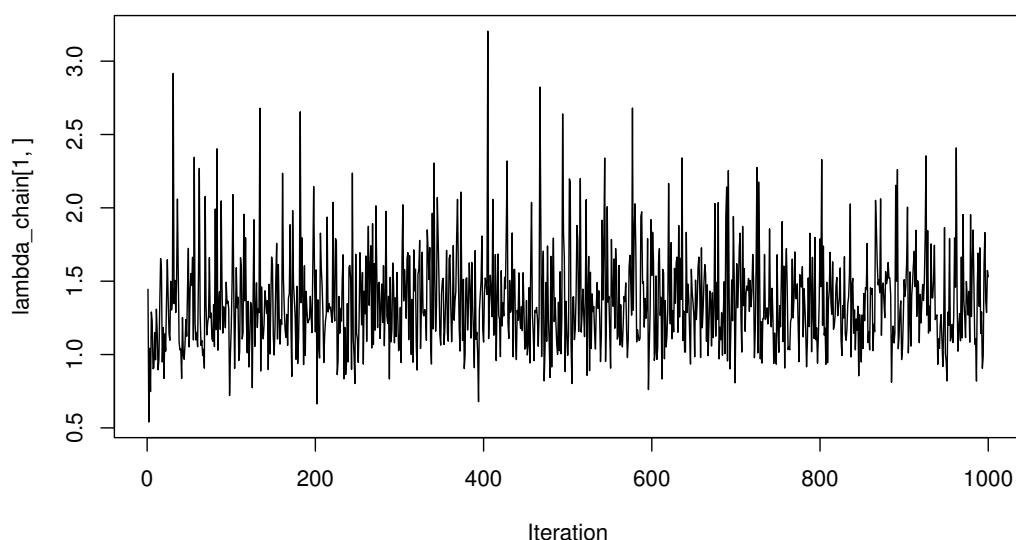


图 3: λ 的马氏链

对于参数 λ ，由于我们预先已经知道，随着迭代的进行，他最终会服从倒伽马分布，而其马氏链在一个区间中震荡也恰恰验证了这个分析，也侧面证明了关于 λ 的马氏链收敛性。

我们可以计算这个马氏链之中各个参数的接受率，如表2所示：

表 2: 各维度 MDSIC 结果

参数	\mathbf{X}	σ^2
接收率	53.2%	50.8%

从表2中可以发现，我们马氏链中关于两个参数 \mathbf{X} 与 σ^2 的接收率都维持在 0.5 以上，能够有效的接受数据。不至于在维度过高时导致接收率过低，算法效率低下，这都是 Metropolis-Hastings 算法的优越性体现。