

基于变分贝叶斯的因子模型参数估计

李亚磊, 陈博文, 李兴平

云南师范大学 数学学院, 昆明 650500

摘要:目的 解决因子分析模型后验分布函数复杂计算的问题。方法 首先利用变分贝叶斯(Variational Bayes, VB)方法,对服从指数分布的因子分析模型进行参数估计,其中 VB 方法使用坐标上升变分推断(Coordinate Ascent Variational Inference, CAVI)算法对参数进行迭代求解,然后与 MCMC(Makov Chain Monte Carlo)方法进行比较,通过随机模拟揭示了当样本量为 300 时,VB 方法和 MCMC 方法的有效性,绘制 ELBO 图判断了 VB 方法的收敛性,绘制参数追踪图、自相关图判断了 MCMC 方法的收敛性,并根据预测值和真实值之间的偏差判断两种方法的好坏,最后对样本量为 713 时的实际数据集进行了实证分析。结果 通过模拟和实证分析可以看到两种方法都是收敛的,偏差绝对值均小于 0.1,但是 VB 方法在估计精度、计算复杂度和运行时间上均优于 MCMC 方法,特别是当样本容量很大时,VB 方法的优势更加明显。结论 在因子分析模型中,因子服从指数分布有时是一种合理的选择,与 MCMC 方法相比,变分贝叶斯方法可以有效减少因子分析模型后验分布函数的计算量,估计参数的效果更好。VB 方法的 3 个优势:第一,由于 VB 方法没有涉及复杂后验分布积分的计算,因此计算上更加简便;第二,由于 VB 方法基于一个近似分布,特别是当样本量很大时,VB 方法的运行时间远远小于 MCMC 方法;第三,从运算结果来看,VB 方法的估计精度要优于 MCMC 方法。

关键词:变分贝叶斯方法;MCMC 方法;因子模型;参数估计

中图分类号: O212.8 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2025.0006.016

Parameter Estimation of Factor Model Using Variational Bayes Approach

LI Yalei, CHEN Bowen, LI Xingping

School of Mathematics, Yunnan Normal University, Kunming 650500, China

Abstract: Objective This study aims to address the challenge of complex calculations for the posterior distribution function in the factor analysis model. **Methods** Initially, the variational Bayes (VB) method is employed to estimate the parameters of the factor analysis model where the factors follow an exponential distribution. The VB method utilizes the coordinate ascent variational inference (CAVI) algorithm for iterative parameter solving. Subsequently, a comparison is made with the MCMC (Makov Chain Monte Carlo) method. Through random simulations, the effectiveness of both the VB and MCMC methods is evaluated when the sample size is 300. The ELBO plot is used to assess the convergence of the VB method, while the parameter trace plot and autocorrelation plot are used to determine the convergence of the MCMC method. The performance of the two methods is judged based on the deviation between the predicted and true values.

收稿日期: 2023-05-12 **修回日期:** 2023-08-19 **文章编号:** 1672-058X(2025)06-0123-12

基金项目: 国家自然科学基金(12261108)。

作者简介: 李亚磊(2000—),男,河北邯郸人,硕士研究生,从事应用统计研究。

通信作者: 李兴平(1976—),男,云南嵩明人,副教授,从事统计计算与推断研究。Email: xingping.li@ynnu.edu.cn.

引用格式: 李亚磊,陈博文,李兴平. 基于变分贝叶斯的因子模型参数估计[J]. 重庆工商大学学报(自然科学版), 2025, 42(6): 123-134.

LI Yalei, CHEN Bowen, LI Xingping. Parameter estimation of factor model using variational Bayes approach[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(6): 123-134.

Finally, an empirical analysis is carried out on an actual dataset with a sample size of 713. **Results** Simulation and empirical analyses indicate that both methods converge, with the absolute deviations being less than 0.1. Nevertheless, the VB method outperforms the MCMC method in terms of estimation accuracy, computational complexity, and running time. The advantages of the VB method become more prominent when dealing with large sample sizes. **Conclusion** In the factor analysis model, assuming that factors follow an exponential distribution can be a reasonable option. Compared to the MCMC method, the variational Bayes method effectively reduces the computational burden associated with the posterior distribution function of the factor analysis model and provides more accurate parameter estimates. The VB method offers three key advantages: First, it simplifies calculations as it avoids the integration of complex posterior distributions. Second, being based on an approximate distribution, the VB method significantly reduces running time, especially for large-scale samples, compared to the MCMC method. Third, the VB method demonstrates higher estimation accuracy than the MCMC method.

Keywords: variational Bayes method; MCMC method; factor model; parameter estimation

1 引言

因子分析作为一种数据降维方法,目前广泛地在贝叶斯统计、信息科学、行为科学、社会科学等领域中使用^[1]。传统因子模型建模优势体现在 3 个方面:首先,能够将多变量的数据维度降低,使研究者能利用高维的数据;其次,能够将可观测的数据转化为一些潜在的或者不可观察的特征因子;最后,作为一种聚类技术,可以使研究者将可观测的数据进行聚类分析^[2]。对因子模型的参数进行估计,常用的方法有 ML(Maximum Likelihood)方法^[3]、EM(Expectation Maximization)算法^[4]以及 MCMC(Markov Chain Monte Carlo)方法^[5]。岳博等^[6]提出两阶段混合因子分析算法,并给出混合因子分析 EM 算法的实现,结果表明了两阶段混合因子分析算法的有效性,但是容易陷入局部最大值。在上述研究中,大多数因子模型的假定建立在经典统计理论的基础之上,没有考虑参数随机化问题,往往与实际不符合^[7]。

由于贝叶斯统计考虑的是参数随机化的问题,引起国内外很多学者的关注。Papastamoulis^[8]使用贝叶斯混合因子分析方法研究了过拟合问题,为因子个数的确定和过拟合推断提供了一种简单、有效的方法;夏业茂等^[9]对于两部分因子分析模型,在贝叶斯框架下,运用 MCMC 方法进行了后验分析,使用 GIBBS 采样器从后验分布中抽取样本,结果表明该方法的有效性和实用性,但是当因子模型的后验分布函数涉及复杂、高维等问题的时候,ML 方法和 EM 算法的求解会变得非常麻烦,而且容易出现过拟合情况,特别是当样本量很大的时候,MCMC 方法的收敛速度就会变慢,与此同时,马尔科夫链的收敛性也难以保证。

为了克服参数估计的计算问题,变分近似方法逐渐被人们接受,这是一种以较少运算量进行参数估计

的方法。Attias^[10]最早提出变分贝叶斯思想,还将其应用到了图像数据模型的参数估计上,变分近似思想由此而来。当利用贝叶斯进行统计推断的时候,一个公认的难题是后验分布函数的计算。廖雪丽等^[11]针对两部分潜变量模型,使用变分贝叶斯方法进行参数估计、变量选择和模型评价等,并与 MCMC 方法进行了对比,结果表明变分贝叶斯方法具有有效性和实用价值;徐定杰等^[12]基于混合高斯分布模型,利用变分贝叶斯方法对参数进行估计,还进行了详细的推导,一定程度上为变分贝叶斯学习算法提供了理论支持;Giordano 等^[13]通过推导一个简单的公式来计算无穷小模型扰动对 MFVB(Mean-field Variational Bayes)后验均值的影响,不足之处是有时会错误地估计方差和协方差;Tzikas 等^[14]在线性回归模型和高斯混合模型的背景下,使用变分贝叶斯方法解决了 EM 算法不能应用于复杂贝叶斯模型的问题,结果表明变分贝叶斯推断方法相较于 EM 算法更有效。

在对因子分析模型的研究中,发现大多数文献基于传统的因子模型,因子被假定为标准正态分布,因子取值可正可负,而实际生活中的数据有时是非负的,因此有必要对因子服从的分布进行推广。当因子服从指数分布时,计算会变得复杂,与 MCMC 方法相比,变分贝叶斯方法可以有效解决计算的问题,但是很少有文献和学者基于变分贝叶斯研究因子服从指数分布的参数估计问题。因此本文利用变分贝叶斯方法和 MCMC 方法对因子服从指数分布的模型进行统计推断并对比不同方法的参数估计效果。

2 理论模型

2.1 因子分析模型

模型的矩阵形式为

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon} \quad (1)$$

其中, $\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_p)^T$ 为观测数据集,均值 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_p)^T$, $\mathbf{A}_{p\times m}=\{a_{ij}\}$ 为因子载荷矩阵, $\mathbf{f}=(\mathbf{f}_1,\mathbf{f}_2,\cdots,\mathbf{f}_m)^T$ 为公共因子,是不可观测的潜变量, $\boldsymbol{\varepsilon}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_p)^T$ 。 $\boldsymbol{\varepsilon}$ 的协方差矩阵记为 Σ , $\boldsymbol{\varepsilon}_i$ 为第 i 个特殊因子,只与 x_i 有关; a_{ij} 为第 i 个变量在第 j 个因子上的载荷,表示第 i 个变量在多大程度上依赖第 j 个因子。当 $\mathbf{x}$ 进行标准化时, a_{ij} 就是第 i 个变量与第 j 个公共因子的相关系数。为了能够计算因子 $\mathbf{f}$,通常假定:

- (1) $\mathbf{f}$ 和 $\boldsymbol{\varepsilon}$ 相互独立且服从高斯分布, $E(\mathbf{f})=0$, $\text{cov}(\mathbf{f})=\mathbf{I}_m$ 。
- (2) $E(\boldsymbol{\varepsilon})=0$, $\text{cov}(\boldsymbol{\varepsilon})=\mathbf{D}$, $\mathbf{D}=\text{diag}(d_1,d_2,\cdots,d_p)$ 。
- (3) 各个公共因子 $\mathbf{f}_i$ 之间和各个特殊因子 $\boldsymbol{\varepsilon}_i$ 之间不相关。

记因子模型的所有未知参数为 $\boldsymbol{\Theta}=(\mathbf{A},\mathbf{f},\mathbf{D})$, 如果将可观测的原始变量 $\mathbf{x}$ 进行 N 次观测,则对应的样本观测值为

$$\mathbf{x}_i=(x_{1i},x_{2i},\cdots,x_{pi})^T,i=1,2,\cdots,N$$

为了简化模型,若样本的观测值已经标准化处理,则对应的样本均值为 0,于是如式(1)所示的分量形式可写为^[6]

$$\mathbf{x}_i=\mathbf{A}\mathbf{f}_i+\boldsymbol{\varepsilon}_i\tag{2}$$

其中, $\mathbf{f}_i=(f_{1i},f_{2i},\cdots,f_{mi})^T$, $\boldsymbol{\varepsilon}_i=(\varepsilon_{1i},\varepsilon_{2i},\cdots,\varepsilon_{pi})^T$ 。

2.2 因子模型的检验

在进行因子分析之前,需要判断数据是否适合进行因子分析,常见的检验方法有 KMO 检验 (Kaiser Meyer Olkin) 和 巴特利特球形检验 (Bartlett's Sphere)^[15]。

KMO 检验最早由 Kaiser、Meyer 和 Olkin 提出,它比较原始变量之间简单相关系数和偏相关系数的相对大小,根据 Kaiser 给出的标准,其值如表 1 所示。

表 1 KMO 值检验

Table 1 KMO value test

KMO 值	标准
>0.9	非常适合
0.8~0.9	适合
0.7~0.8	一般
0.6~0.7	不太适合
<0.5	不适合

巴特利特球形检验常用来检验各个变量之间的相关程度,并判断是否适合做因子分析。如果该值较大,且其对应的 p 值小于显著性水平 (一般为 0.05),则应

该拒绝原假设,适合做因子分析。

2.3 因子个数的选择

在估计公共因子和因子载荷矩阵之前,需要确定公共因子 K 的个数^[16]。常用的方法包括似然比检验、平行分析以及碎石图等。这些常用方法的核心是考虑前 K 个特征值所能解释的方差占比,公式为

$$\hat{p}_k=\frac{\sum_{j=1}^k\lambda_j(\hat{\Sigma})}{\sum_{j=1}^p\lambda_j(\hat{\Sigma})}$$

若选到某个 K 之后, $\hat{p}_k$ 就不再显著地增加,则此时的 K 值为确定的因子个数,本文主要使用这种方法来进行。选择因子个数的方法还有许多,例如基于相邻特征值之差的因子个数估计方法^[17]。

2.4 最大后验估计

假设 $\mathbf{y}$ 是观测向量,而 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ 是待估未知参数向量,参数的后验分布为

$$p(\boldsymbol{\theta}|\mathbf{y})=\frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

这里, $p(\mathbf{y})=\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ 。当数据 $\mathbf{y}$ 是 n 维时, $\boldsymbol{\theta}$ 的 MAP 估计为

$$\hat{\boldsymbol{\theta}}_{\text{MAP}}=\arg\max_{\boldsymbol{\theta}}p(\boldsymbol{\theta}|\mathbf{y})=\arg\max_{\boldsymbol{\theta}}p(\mathbf{y}|\boldsymbol{\theta})\cdot p(\boldsymbol{\theta})$$

由于分子中的 $p(\mathbf{y})$ 相对于 $\boldsymbol{\theta}$ 是独立的,所以在求 $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ 时可以直接忽略掉 $p(\mathbf{y})$,对结果没有影响。为了得到满足条件的 $\boldsymbol{\theta}$,需要对 $p(\boldsymbol{\theta}|\mathbf{y})$ 中的 $\boldsymbol{\theta}$ 求导,并令其等于 0,即

$$\frac{p(\boldsymbol{\theta}|\mathbf{y})}{\partial\boldsymbol{\theta}}=\frac{p(\mathbf{y}|\boldsymbol{\theta})\cdot p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=0$$

这时的 $\boldsymbol{\theta}$ 就称为最大后验估计^[18],记作 $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ 。

2.5 MCMC 方法

MCMC 方法是一种使用随机抽样来近似复杂积分的方法,该方法包括马尔科夫链和蒙特卡罗两部分。MC (Monte Carlo) 积分^[5]是利用有关分布的随机抽样来近似积分的一种方法,但从复杂概率分布抽样很困难或者无法直接做到时,就需要马尔科夫链。马尔科夫链是一种序贯模型,它依概率的方式从一种状态转移到另一种状态,其中链所采取的下一个状态取决于以前的状态,即所谓的马尔科夫性。如果马尔科夫链构造得当并运行很长时间,那么它也将从目标概率分布中提取样本的状态。通过构造一个马尔科夫链,从相应的积分中抽样,然后利用蒙特卡罗积分来近似积

分。随着 MCMC 方法的不断发展,一个比较好的 MCMC 方法是 NUTS 方法,它是在 HMC 方法的基础上发展而来的,下面对两种方法进行简单介绍。

HMC(Hamiltonian Monte Carlo)^[19]是通过哈密顿动力学来构造的 MCMC,也称为混合蒙特卡罗方法,可以归类为辅助变量 MCMC 方法之一。因为它结合了 MCMC 和确定性模拟方法,可以实现哈密顿动力学的越级方法。无掉头抽样器(No-U-Turn Sampler, NUTS)^[20]是 HMC 的扩展,是为了消除设置 HMC 越级数量的需要。NUTS 是目前比较好的一种 MCMC 方法,本文抽样中使用的是这种方法。

3 变分贝叶斯参数估计

3.1 参数的先验分布

在贝叶斯框架下,对参数先验分布设置常用的方法有共轭分布法、Jefferys 原则、最大熵原则等。对于式(2),记 $\theta = (A, f, D)$ 是未知参数,由于特殊因子服从多维高斯分布,于是可将 $x_i (i=1, 2, \dots, N)$ 的概率密度表示为

$$p(x_i | A, f_i, D) = N(x_i | Af_i, D)$$

这里 $p(\cdot)$ 表示概率密度函数, $D = \sigma^2 I$, 则参数 $\theta = (A, f, D)$ 的联合似然函数表示为

$$p(x | A, f, D) = \prod_{i=1}^N N(x_i | Af_i, D)$$

当确定因子的个数后,根据式(1)的假设可以知道因子的先验服从 $N(0, I_m)$ 的多元高斯分布,取值可正可负。但是在实际生活中有时会有非负数据,这时假设因子的先验服从指数分布是一个合理的选择^[21]。本文选取因子的先验分布为参数服从 $\beta_k (k$ 为因子的个数) 的指数分布。

对于因子载荷矩阵 A 的先验设置,假设 A 的每一列服从均值为 μ 和不同精度 v 的高斯分布^[10]:

$$p(A | v) = \prod_{j=1}^q p(A_{\cdot j} | v_j) = \prod_{j=1}^q N\left(A_{\cdot j} | \mu, \frac{I_p}{v_j}\right)$$

其中, $A_{\cdot j}$ 表示矩阵 A 的第 j 列, $v = \{v_1, v_2, \dots, v_m\}$ 。

由于参数 D 不随组件维数的增加而增加,所以并没有给出参数 D 的先验,而是把参数 D 当作超参数来处理^[22]。

3.2 变分贝叶斯参数估计

设可观测数据 $x = (x_1, x_2, \dots, x_p)$, 参数 $\theta = \{\theta_1, \theta_2, \dots, \theta_m\}$, 根据贝叶斯公式

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{\int p(x | \theta)p(\theta)d\theta}$$

其中, $p(\theta | x)$ 称为后验分布, $p(\theta)$ 称为先验分布, $p(x | \theta)$ 称为概率函数。

对于一个给定的先验分布 $p(\theta)$, 目标是求解它的后验分布 $p(\theta | x)$ 。目前经典参数估计的方法有极大似然法(ML)^[3]、期望最大法(EM)^[4]、马尔科夫链蒙特卡罗法(MCMC)^[5]等。ML 算法是通过极大化联合似然函数对参数进行估计,如果参数的个数非常多,联合似然函数的求解优化就会变得非常复杂;EM 算法是通过添加数据的思想,建立一个隐变量模型,然后进行 ML 估计,算法流程是给定参数的一个初始分布,进而不断进行迭代,有效降低了 ML 的复杂度,但是和 ML 一样,容易出现数据过拟合的情况;MCMC 方法的思想是从给定的分布中抽取满足条件的样本,给定一个马尔科夫链,使其满足细致平衡条件,这样可以有效避免后验分布的复杂积分计算,但是计算量大,运行时间长,容易出现抽样偏差。

为了解决上述问题,本文在传统贝叶斯推断和 EM 迭代算法的基础上,利用变分贝叶斯方法对未知参数进行估计。该方法通过引入一种变分近似理论,用一个较易处理的分布 $q(\theta)$ 来逼近参数真实的后验分布 $q(\theta | x)$, 这样就有效避免了后验分布的复杂计算。

定义 1 KL(Kullback Leibler)散度:假设可观测数据 $x = (x_1, x_2, \dots, x_p)$, M 维的未知参数向量记为 $\theta \in R^m$, 则未知参数的后验分布为 $p(\theta | x)$, 记近似分布为 $q(\theta) \in Q$, Q 为一个分布族,用 KL 散度来衡量两个分布之间的距离

$$D_{KL}(q(\theta) || p(\theta | x)) = - \int \ln\left(\frac{p(\theta | x)}{q(\theta)}\right) q(\theta) d\theta$$

根据恒等式

$$\ln p(x) = \ln \frac{p(x, \theta)}{q(\theta)} - \ln \frac{p(\theta | x)}{q(\theta)} \quad (3)$$

对式(3)两边同乘以 $q(\theta)$, 对 θ 进行积分

$$\begin{aligned} \ln p(x) &= \int q(\theta) \ln \frac{p(x, \theta)}{q(\theta)} d\theta - \int q(\theta) \ln \frac{p(\theta | x)}{q(\theta)} d\theta = \\ &L(q) + D_{KL}(q(\theta) || p(\theta | x)) \end{aligned}$$

其中,

$$L(q) = \int q(\theta) \ln \frac{p(x, \theta)}{q(\theta)} d\theta, \int q(\theta) d\theta = 1$$

由于 KL 散度距离是非负的,故

$$\ln p(x) \geq L(q)$$

将寻找到的最优近似分布记为 $q^*(\theta)$, 即

$$q^*(\boldsymbol{\theta}) = \operatorname{argmin}_{q(\boldsymbol{\theta}) \in \mathcal{Q}} D_{\text{KL}}(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta}|\mathbf{x})) = p(\boldsymbol{\theta}|\mathbf{x}) \quad (4)$$

对于可观测数据 $\mathbf{x}$, $\ln p(\mathbf{x})$ 为一个常量, $L(q)$ 称为变分下界(有的称之为证据下界), 于是求解 KL 散度距离最小问题可以转化为求解最大化变分下界。这样式(4)就转化为

$$q^*(\boldsymbol{\theta}) = \operatorname{argmin}_{q(\boldsymbol{\theta}) \in \mathcal{Q}} D_{\text{KL}}(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta}|\mathbf{x})) = \operatorname{argmax}_{q(\boldsymbol{\theta}) \in \mathcal{Q}} L(q) = p(\boldsymbol{\theta}|\mathbf{x})$$

当固定参数的初始值后, 使用坐标上升变分推断(CAVI)算法(也称为 VB 方法)进行迭代求解, 可以得到更新的公式:

$$q_j^*(\boldsymbol{\theta}_j) = \operatorname{argmax}_{q_j} L(q) = \frac{\exp\{E_{q(\boldsymbol{\theta}_{k \neq j})}[\ln p(\boldsymbol{\theta}, \mathbf{x})]\}}{\int \exp\{E_{q(\boldsymbol{\theta}_{k \neq j})}[\ln p(\boldsymbol{\theta}, \mathbf{x})]\} d\boldsymbol{\theta}_j}$$

根据平均场理论^[23], 变分分布可以将参数和潜在变量进行因式分解, 比如将 $\boldsymbol{\theta}$ 划分为 $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_m$, 假设每个参数之间是相互独立的, 则有

$$q(\boldsymbol{\theta}) = \prod_{j=1}^m q(\boldsymbol{\theta}_j | \mathbf{x})$$

虽然能够对每一个参数进行一次划分, 但是变分分布会变得比较复杂。有时将几个参数放在一起会更容易处理, 于是本文选择将几个参数放在一起。VB (Variable Bayes) 方法的具体实现过程如表 2 所示。

表 2 VB 方法

Table 2 Variational Bayes method

Algorithm 1 VB 方法
输入: 给定模型 $p(\mathbf{x}, \boldsymbol{\theta})$, 数据集 $\mathbf{x}$ 。
输出: 平均值场领域的变分密度。
$q(\boldsymbol{\theta}) = \prod_{i=1}^N q_i(\boldsymbol{\theta}_i)$
1. 初始化 $q_i(\boldsymbol{\theta}_i)$, $i=1, 2, \dots, N$ 。
2. 当变分上界不收敛时, 开始循环:
对于 $i \in \{1, 2, \dots, N\}$,
令 $q_i^*(\boldsymbol{\theta}_i) = \frac{\exp\{E_{q(\boldsymbol{\theta}_{k \neq i})}[\ln p(\boldsymbol{\theta}, \mathbf{x})]\}}{\int \exp\{E_{q(\boldsymbol{\theta}_{k \neq i})}[\ln p(\boldsymbol{\theta}, \mathbf{x})]\} d\boldsymbol{\theta}_i}$
终止于:
变分上界 $L(q) \leq t$, t 为给定的阈值,
$L(q) = \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{x}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}$
3. 结束循环。
4. 返回 $q(\boldsymbol{\theta})$ 的值。

3.3 变分后验分布

对于模型式(2), 记参数 $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{f}, \mathbf{D})$, 由于参数 $\mathbf{D}$ 是一个超参数, 故没有给定先验, 根据 VB 方法, 接下来主要讨论参数 $(\mathbf{A}, \mathbf{f}_i)$, $i=1, 2, \dots, N$ 的变分后验分布。

$q(\mathbf{A})$ 的变分后验分布:

$$\ln(q^*(\mathbf{A})) = \frac{E_{f_j}[\ln \prod_{i=1}^N p(\mathbf{y}_i | \mathbf{A}, \mathbf{x}_i, \mathbf{D})]}{\int E_{f_j}[\ln \prod_{i=1}^N p(\mathbf{y}_i | \mathbf{A}, \mathbf{x}_i, \mathbf{D})] d\mathbf{A}}$$

$q(\mathbf{f}_j)$ 的变分后验分布:

$$\ln(q^*(\mathbf{f}_j)) = \frac{E_{\mathbf{A}}[\ln \prod_{i=1}^N p(\mathbf{y}_i | \mathbf{A}, \mathbf{x}_i, \mathbf{D})]}{\int E_{\mathbf{A}}[\ln \prod_{i=1}^N p(\mathbf{y}_i | \mathbf{A}, \mathbf{x}_i, \mathbf{D})] d\mathbf{f}_j}$$

本节主要研究使用 VB 方法估计参数的相合性和渐进正态性。参考 YOU 等^[24]的假设, 首先给出以下正则条件:

A1 对于 $1 \leq i \leq n$, $\mathbf{x}_i = \mathbf{A}\mathbf{f}_i + \boldsymbol{\varepsilon}_i$, 其中 $\boldsymbol{\varepsilon}_i$ 是独立同分布的, 服从 $N(0, \sigma_0^2)$, 并且 $\mathbf{A}_0$ 和 $0 \leq \sigma_0^2 \leq \infty$ 分别是 $\mathbf{A}$ 和 σ^2 的真实值, $\mathbf{A}_0$ 的每个元素为有限值。

A2 对于 $1 \leq i \leq n$, 随机向量 $\mathbf{f}_i \in \mathbf{R}^m$ 独立同分布, m 是固定的 ($m < n$)。

A3 $m \times m$ 矩阵 $\mathbf{E}(\mathbf{f}_i \mathbf{f}_i^T)$ 每个元素有限, 并且 $\mathbf{f} = [\mathbf{f}_1, \dots, \mathbf{f}_m]$, 其中对于 $n \rightarrow \infty$, 有 $P(\mathbf{a}^T \mathbf{f} = 0) < 1, \forall \mathbf{a} \neq 0 \Rightarrow P(\text{rank}(\mathbf{f}) = m) \rightarrow 1$ 。

A4 对于 $1 \leq i \leq n$, 随机向量 $\mathbf{f}_i$ 和 $\boldsymbol{\varepsilon}_i$ 相互独立。

令 $\mathbf{U}(\text{diag}(\boldsymbol{\lambda})) \mathbf{U}^T$ 为 $\mathbf{f}^T \mathbf{f}$ 的特征值分解, 其中 $\mathbf{U}$ 为正交矩阵, 并且 $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^T$ 是特征值向量, 对于 $i=1, \dots, m, \lambda_i > 0$, 同样令

$$\begin{aligned} \mathbf{A}_n &= \mathbf{n}^{-1} \mathbf{f}^T \mathbf{f} \\ \mathbf{b}_n &= \mathbf{n}^{-1} \mathbf{f}^T \mathbf{x} \\ c_n &= \text{tr}(\mathbf{A}_n (\mathbf{A}_n + \sigma_\beta^{-2} \mathbf{n}^{-1} \mathbf{d}_n \mathbf{I})^{-1}) \\ \mathbf{d}_n &= \mathbf{B}_{q(\sigma^2)} / (n/2) \end{aligned}$$

定理 1 假设满足 A1—A4 的正则性条件下, 变分贝叶斯方法的参数 $\mathbf{A}$ 满足渐进正态性。

证明 变分密度 $q(\boldsymbol{\theta})$ 对应的最优密度形式为 $q^*(\mathbf{A}) \sim N(\boldsymbol{\mu}_{q(\mathbf{A})}, \boldsymbol{\Sigma}_{q(\mathbf{A})})$, 假设 $\boldsymbol{\mu}_{q(\mathbf{A})}$ 、 $\boldsymbol{\Sigma}_{q(\mathbf{A})}$ 是变分密度的最佳参数。

(1) 证明 $\mathbf{A}_{\text{VB}}$ 是 $\mathbf{A}_0$ 的一致估计量。

$\mathbf{A}_{\text{VB}}$ 可以写成 $\mathbf{A}_{\text{VB}} = (\mathbf{A}_n + \sigma_\beta^{-2} \mathbf{n}^{-1} \mathbf{d}_n \mathbf{I})^{-1} \mathbf{b}_n$, 在 A1—

A4 的假设下, $d_n = O_p(1)$ 并且 $d_n^{-1} = O_p(1)$, 所以项 $\sigma_\beta^{-2} n^{-1} d_n$ 是 $O_p(n^{-1})$, 因此可以忽略不计。

由于几乎肯定的收敛意味着概率上的收敛, 并且

$\mu_{q(A)} = A_{VB}$, 于是有 $A_{VB} \xrightarrow{P} [E(f f_i^T)]^{-1} E(f f_i^T) A_0 = A_0$ 。故 A_{VB} 是 A_0 的一致估计量。

(2) 证明 $\Sigma_{q(A)}$ 是 $\sigma_{unbiased}^2 (f^T f)^{-1}$ 的一致估计量。

$$\Sigma_{q(A)} = \left(\frac{\|x - \hat{A}_{LS} f\|}{n-p} \right) (f^T f)^{-1} + O(\sigma_A^{-2})$$

对于 A_0 和 σ_0^2 , 常用的无偏估计是 $A_{LS} = (f^T f)^{-1} f^T x$, $\sigma_{unbiased}^2 = \|x - A_{LS} f\|^2 / (n-p)$, 注意当 $\sigma_A^2 \rightarrow \infty$ 时, 有 $A_{VB} = A_{LS} + O(\sigma_A^{-2})$, 并且

$$\sigma_{VB}^2 = \frac{(n-p) \sigma_{unbiased}^2}{n-p-2(1-\frac{p}{n})} + O(\sigma_A^{-2})$$

于是当 $\sigma_A^2 \rightarrow \infty$ 和 $n \rightarrow \infty$ 时, σ_{VB}^2 接近 $\sigma_{unbiased}^2$ 。故当 $\sigma_A^2 \rightarrow \infty$ 时, $\Sigma_{q(A)}$ 接近 $\sigma_{unbiased}^2 (f^T f)^{-1}$, 所以 $\Sigma_{q(A)}$ 是 $\sigma_{unbiased}^2 (f^T f)^{-1}$ 的一致估计量。

所以变分贝叶斯方法的参数 A 依分布收敛到正态分布, 期望为 A_0 , 渐进方差为 $\sigma_{unbiased}^2 (f^T f)^{-1}$, 满足渐进正态性。

定理 2 设 $\tilde{\Theta}^n$ 为参数 Θ 的极大似然估计, 具有强相合性, $\hat{\Theta}^n$ 表示变分贝叶斯估计, 则变分贝叶斯以 $O(1/n)$ 的速率局部收敛于极大似然估^[25]。

证明 假设对于 $s=1, \dots, m$, $\tilde{\Theta}^n$ 是下列似然方程的解:

$$L_s^n(\Theta) \stackrel{\Delta}{=} \pi_s - \frac{1}{n} \sum_{i=1}^n \frac{p_s(y_i) \pi_s}{p(y_i)} = 0$$

$$\bar{L}_s^n(\Theta) \stackrel{\Delta}{=} \mu_s - \left\{ \frac{1}{n} \sum_{i=1}^n y_i \frac{p_s(y_i)}{p(y_i)} \right\} / \left\{ \frac{1}{n} \sum_{i=1}^n \frac{p_s(y_i)}{p(y_i)} \right\} = 0$$

$$\tilde{L}_s^n(\Theta) \stackrel{\Delta}{=} \Gamma_s - \left\{ \frac{1}{n} \sum_{i=1}^n \frac{p_s(y_i)}{p(y_i)} \right\} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{p_s(y_i)}{p(y_i)} \right\} \times$$

$$(y_i - \mu_s)(y_i - \mu_s)' \}^{-1} = 0$$

$\hat{\Theta}^n$ 表示变分贝叶斯估计, 是迭代算法。

$$\Theta^{(k)} = (1-\varepsilon) \Theta^{(k-1)} + \varepsilon \begin{pmatrix} \Pi(\Theta^{(k-1)}) \\ M(\Theta^{(k-1)}) \\ S(\Theta^{(k-1)}) \end{pmatrix} \stackrel{\Delta}{=}$$

$$\Phi_n^e((\Theta^{(k-1)}))$$

在真实值领域的稳定点, 有 $\hat{\Theta}^n$ 满足

$$\hat{\Theta}^n - \begin{pmatrix} \Pi(\hat{\Theta}^n) \\ M(\hat{\Theta}^n) \\ S(\hat{\Theta}^n) \end{pmatrix} = 0 \quad (5)$$

并且变分后验分布 $q(\pi)$ 、 $q(\Gamma_s)$ 和 $q(\mu_s | \Gamma_s)$ 的超参数为

$$\hat{\lambda}_s^n = n \hat{\pi}_s^n + \lambda^0$$

$$\hat{\rho}_s^n = (n \hat{\mu}_s^n \hat{\pi}_s^n + \beta^0 \rho^0) / (n \hat{\pi}_s^n + \beta^0)$$

$$\hat{\beta}_s^n = n \hat{\pi}_s^n + \beta^0, \hat{v}_s^n = n \hat{\pi}_s^n + v^0$$

$$\hat{\Theta}_s^n = n \hat{\pi}_s^n (\hat{\Gamma}_s^n)^{-1} + n \hat{\pi}_s^n \beta^0 (\hat{\mu}_s^n - \rho^0) \times$$

$$(\hat{\mu}_s^n - \rho^0)' (n \hat{\pi}_s^n + \beta^0)^{-1} + \Phi^0$$

当 $k \rightarrow \infty$ 时, $\hat{\gamma}_{is}^{(k)}$ 、 $\hat{\gamma}_i^{(k)}$ 表示 $\gamma_{is}^{(k)}$ 、 $\gamma_i^{(k)}$ 的收敛值, $\hat{p}_{is}^n$ 、 $\hat{p}_i^n$ 表示 p_{is}^n 、 p_i^n 在 $\hat{\Theta}^n$ 点的值, 从式(5)的第一个方程可以得到

$$0 = \hat{\pi}_s^n - \frac{1}{n} \sum_{i=1}^n \frac{\hat{\gamma}_{is}^n}{\hat{\gamma}_i^n} =$$

$$L_s^n(\hat{\Theta}^n) + \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\hat{\pi}_s^n \hat{p}_{is}^n}{\hat{p}_i^n} - \frac{\hat{\gamma}_{is}^n}{\hat{\gamma}_i^n} \right\} =$$

$$L_s^n(\hat{\Theta}^n) + \frac{1}{n} \sum_{i=1}^n \frac{\hat{\pi}_s^n \hat{p}_{is}^n \hat{\gamma}_i^n - \hat{p}_i^n \hat{\gamma}_{is}^n}{\hat{p}_i^n \hat{\gamma}_i^n} \quad (6)$$

根据 γ_{is} 的定义, 可以得到

$$\gamma_{is} = \pi_s | \Gamma_s |^{1/2} e^{-(y_i - \mu_s)' \Gamma_s (y_i - \mu_s)/2} + O\left(\frac{1}{n}\right)$$

因此 $\hat{\gamma}_i^n = \hat{p}_i^n + O(1/n)$, 所以式(6)的第二项为 $O(1/n)$ 阶。根据泰勒展开式, 第一项可以重新写为

$$L_s^n(\hat{\Theta}^n) = L_s^n(\tilde{\Theta}^n) + \nabla L_s^n(\tilde{\Theta}^n + \lambda(\hat{\Theta}^n - \tilde{\Theta}^n))(\hat{\Theta}^n - \tilde{\Theta}^n) =$$

$$\nabla L_s^n(\tilde{\Theta}^n + \lambda(\hat{\Theta}^n - \tilde{\Theta}^n))(\hat{\Theta}^n - \tilde{\Theta}^n)$$

其中, $0 \leq \lambda \leq 1$ 。因此, 可以得到

$$0 = \nabla L_s^n(\tilde{\Theta}^n + \lambda(\hat{\Theta}^n - \tilde{\Theta}^n))(\hat{\Theta}^n - \tilde{\Theta}^n) + O(1/n)$$

按照相同的思路, 根据式(5)的第二和第三个方程, 有

$$0 = \nabla \bar{L}_s^n(\tilde{\Theta}^n + \lambda(\hat{\Theta}^n - \tilde{\Theta}^n))(\hat{\Theta}^n - \tilde{\Theta}^n) + O(1/n)$$

$$0 = \nabla \tilde{L}_s^n(\tilde{\Theta}^n + \lambda(\hat{\Theta}^n - \tilde{\Theta}^n))(\hat{\Theta}^n - \tilde{\Theta}^n) + O(1/n)$$

如果, 令

$$L = \begin{pmatrix} L_1^n & \dots & L_m^n & \bar{L}_1^n & \dots & \bar{L}_m^n & \tilde{L}_1^n & \dots & \tilde{L}_m^n \end{pmatrix}^T$$

第三个方程有

$$\nabla L^n(\tilde{\Theta}^n+\lambda(\hat{\Theta}^n-\tilde{\Theta}^n))(\hat{\Theta}^n-\tilde{\Theta}^n)+O(1/n)$$

可见 $\hat{\Theta}^n$ 收敛于真实值 Θ^* ,而已知极大似然估计 $\tilde{\Theta}^n$ 趋近于 Θ^* ,所以变分贝叶斯以 $O(1/n)$ 的速率局部收敛于极大似然估计。

4 数值模拟

本节通过数值模拟,使用变分贝叶斯方法和 MCMC 方法对因子模型的参数进行估计,在数值模拟中选择变量的个数 $p=5$,公共因子的个数为 $k=2$,划分参数时将因子载荷矩阵 A 中的每一行参数放在一起,其中每一行包括 2 个参数,给定的因子载荷矩阵 A 和 D 值为

$$A^T=\begin{bmatrix}0.819 & 0.521 & 0.476 & 0.674 & 0.771 \\ 0.454 & 0.571 & 0.751 & 0.552 & 0.426\end{bmatrix}$$
$$D=\text{diag}(0.124 \quad 0.402 \quad 0.210 \quad 0.241 \quad 0.225)$$

假设因子服从一个参数为 β_k 的指数分布,其中 $\beta_1=14,\beta_2=15$,运用式(2)生成一个 $N=300$ 的数据集,即可观测数据集。

模拟的思路如下:利用给定的参数值生成一个 $N=300$ 的可观测数据集,首先对数据进行标准化,然后运用 MCMC 和 VB 方法进行因子载荷矩阵的参数估计,并且对两种算法进行分析和比较。

表 3 给出了当 $N=300$ 时,MCMC 方法和 VB 方法对于因子模型参数的估计值及偏差,可以看到两种方法的偏差均小于 0.1,说明了两种方法的有效性。在大多数情况下,VB 方法的偏差要小于 MCMC 方法,说明了 VB 方法的估计精度要优于 MCMC 方法。且 MCMC 方法在计算上涉及复杂的积分运算,而 VB 方法没有涉及,因此 VB 方法在计算上更为简单。为了简便,下面给出算法收敛的第一组参数图形。

表 3 两种方法对因子模型参数的估计值及偏差 ($N=300$)

Table 3 Estimated values and deviations of factor model parameters using two methods ($N=300$)

参 数	真实值	MCMC 估计值	偏 差	VB 估 计值	偏 差
A11	0.819	0.809	0.010	0.810	0.009
A12	0.454	0.365	0.089	0.380	0.074
A21	0.521	0.509	0.012	0.510	0.011
A22	0.571	0.513	0.058	0.520	0.051
A31	0.476	0.460	0.016	0.450	0.026
A32	0.751	0.699	0.052	0.730	0.021

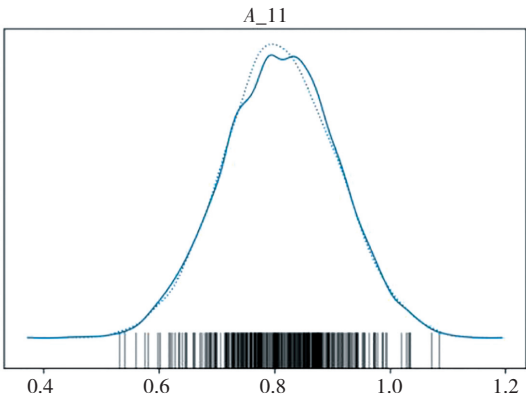
续表(表3)

参 数	真实值	MCMC 估计值	偏 差	VB 估 计值	偏 差
A41	0.674	0.661	0.013	0.660	0.014
A42	0.552	0.480	0.072	0.500	0.052
A51	0.770	0.759	0.011	0.740	0.030
A52	0.426	0.363	0.063	0.370	0.056
D11	0.124	0.212	-0.089	0.200	-0.076
D22	0.402	0.478	-0.075	0.470	-0.067
D33	0.210	0.300	-0.090	0.265	-0.055
D44	0.241	0.333	-0.092	0.314	-0.073
D55	0.225	0.292	-0.067	0.316	-0.090

图 1 给出了当 $N=300$ 时,MCMC 方法对于因子模型参数的追踪图,目的是检验 MCMC 方法的收敛性。图 1(a)、图 1(c)表示随机变量核密度估计(Kernel Density Estimation,KDE)图,是对边际后验直方图进行平滑处理得到的。图中的两条曲线不仅是平滑的曲线,而且都趋近于高斯分布,根据中心极限定理,说明 KDE 图表现良好。图 1(b)、图 1(d)表示随机变量的路径图,是在每一步采样过程中得到的采样值,所有得到的采样值都在没有强周期性和(特别是)一定趋势的区域内,说明路径图表现良好。

图 2 给出了当 $N=300$ 时,MCMC 方法的自相关图,目的也是判断 MCMC 方法的收敛性。可以看到所有生成的采样值之间没有很强的相关性,说明了自相关图表现良好。结合路径图和自相关图,可以判断 MCMC 方法是收敛的,从而保证了所有得到的采样值来源于正确的目标分布。

图 3 给出了当样本量 $N=300$ 时,VB 方法的 ELBO 图,目的是检验 VB 方法的收敛性。从图中可以看到,随着迭代次数不断增加,变分下界 ELBO 在不断地收敛,可以认为 VB 方法是收敛的。



(a) 第一条链 A11 核密度估计图

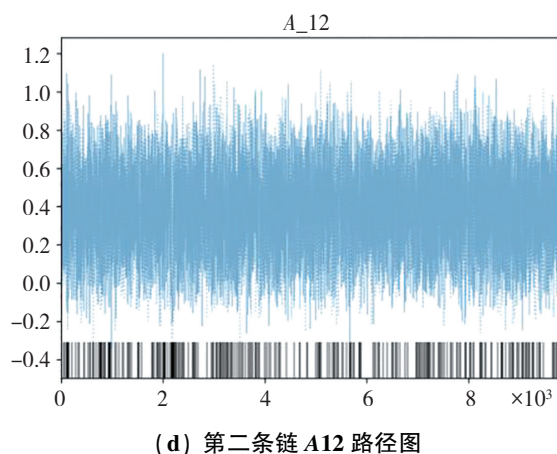
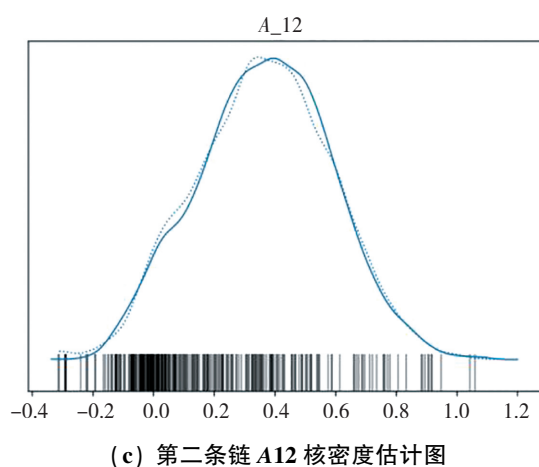
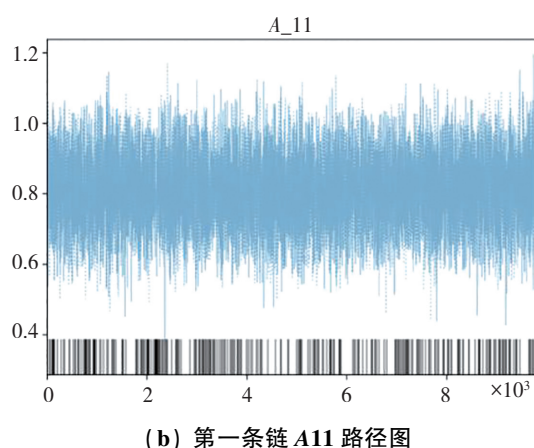
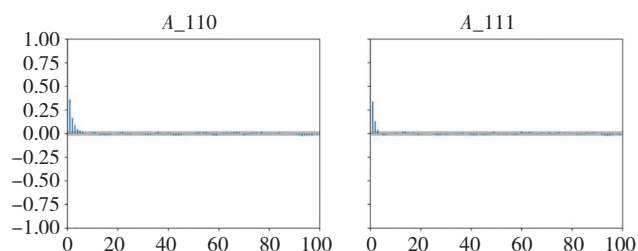
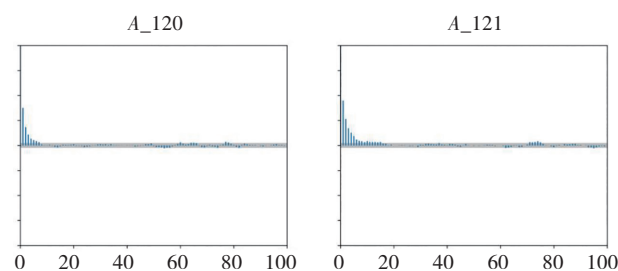
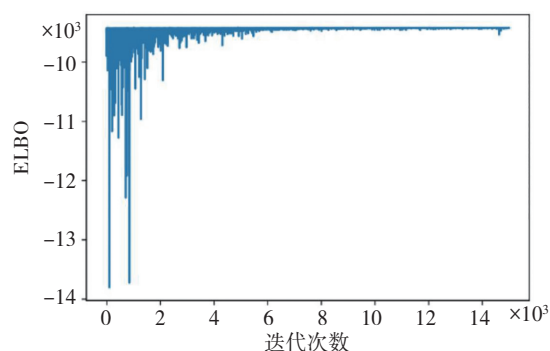
图 1 MCMC 方法的参数追踪图 ($N=300$)Fig. 1 Parameter Trace Plot of MCMC Method ($N=300$)(b) 第二条链 A₁₁ 自相关图(d) 第二条链 A₁₂ 自相关图图 2 MCMC 方法的自相关图 ($N=300$)Fig. 2 Autocorrelation plot of MCMC method ($N=300$)图 3 VB 方法的 ELBO 图 ($N=300$)Fig. 3 ELBO plot of VB method ($N=300$)

表 4 给出了 MCMC 方法和 VB 方法的运行时间,从表中很容易可以看出:VB 方法运行的时间要比 MCMC 方法短,这是 VB 方法的优势之一。特别地,当随着样本量不断增加,这一时间差距会更加明显,VB 方法能节省更多时间。

表 4 MCMC 方法和 VB 方法的运行时间 ($N=300$)Table 4 Running times of MCMC method and VB method ($N=300$)

参 数	MCMC 方法运行时间/s	VB 方法运行时间/s
A1	56	6.29
A2	61	4.64
A3	54	6.35
A4	56	6.07
A5	56	6.13

5 实例分析与应用

在本节中,为了验证本文所提出的当因子服从指数分布时,因子模型的变分贝叶斯方法,选取我国 1999—2021 年的分地区农村居民人均消费支出构成数据集,该数据集可以在国家统计局官网 <http://www.stats.gov.cn/sj/ndsj/2022/indexch.htm> 中找到。从这

个数据集中,选取不同地区的农村居民在饮食、穿衣、住房、日用品、行走方面的人均消费支出共 5 个变量,样本量 $N=713$,划分参数时同样将因子载荷矩阵 A 中的每一行参数放在一起,其中每一行包括 2 个参数。

实例分析的主要步骤是:首先进行 KMO 检验和巴特利特球形检验,判断数据是否适合进行因子分析,如果适合就确定因子个数,然后使用 VB 方法进行因子模型的参数估计,并与 MCMC 方法进行对比。

5.1 两种检验

由表 5 可知:当样本量 $N=713$ 时,通过了 KMO 和巴特利特球形检验,说明数据集适合进行因子分析。

表 5 KMO 检验和巴特利特球形检验的结果 ($N=713$)

Table 5 Results of KMO test and Bartlett spherical test ($N=713$)

样本量 N	KMO 检验	巴特利特球形检验
713	0.89	0

5.2 因子个数的选择

由表 6 可知,当样本量 $N=713$ 时,因子个数选取 2 个因子比较合理,因为样本量的特征值累加贡献率达到了 95%以上。

表 6 数据标准化后的特征值累加贡献率 ($N=713$)

Table 6 Cumulative contribution rates of eigenvalues after data standardization ($N=713$)

样本量 N	数据标准化后的特征值累加贡献率				
713	0.93	0.97	0.98	0.99	1

5.3 因子载荷矩阵 A 和 D 的估计

由于变分贝叶斯方法对于参数初始值并不敏感,所以设置的初始值如下:载荷矩阵 A 的精度 $v_1=10, v_2=5$,不可观测因子的参数 $\beta_1=14, \beta_2=15$,因子载荷矩阵 A 和 D 的迭代初值是采用主成分分析法得到的估计值。

$$A^T = \begin{bmatrix} 0.801 & 0.548 & 0.835 & 0.789 & 0.738 \\ 0.510 & 0.834 & 0.508 & 0.572 & 0.632 \end{bmatrix}$$
$$D = \text{diag}(0.099 \quad 0.004 \quad 0.045 \quad 0.050 \quad 0.056)$$

由表 7、表 8 可知在样本量 $N=713$ 时,MCMC 方法和 VB 方法对因子模型参数的估计值及偏差。两种方法偏差的绝对值均小于 0.1,说明了两种方法的有效性,但是两种方法估计的差距不是很大。由于 VB 方法

没有涉及复杂的积分运算,因而在运算上比 MCMC 方法更为简单。

表 7 MCMC 方法对因子模型参数的估计值及偏差 ($N=713$)

Table 7 Estimated values and deviations of factor model parameters by MCMC method ($N=713$)

参 数	主成分法参数值	MCMC 估计值	偏 差
A11	0.801	0.787	0.014
A12	0.510	0.502	0.008
A21	0.548	0.532	0.016
A22	0.834	0.829	0.005
A31	0.835	0.787	0.048
A32	0.508	0.518	-0.010
A41	0.789	0.788	0.001
A42	0.572	0.591	-0.019
A51	0.738	0.727	0.011
A52	0.632	0.603	0.029
D11	0.099	0.129	-0.030
D22	0.004	0.030	-0.025
D33	0.045	0.112	-0.068
D44	0.050	0.030	0.020
D55	0.056	0.108	-0.052

表 8 VB 方法对因子模型参数的估计值及偏差 ($N=713$)

Table 8 Estimated values and deviations of factor model parameters by VB method ($N=713$)

参 数	主成分法参数值	VB 估计值	偏 差
A11	0.801	0.780	0.021
A12	0.510	0.520	-0.010
A21	0.548	0.520	0.028
A22	0.834	0.794	0.040
A31	0.835	0.829	0.006
A32	0.508	0.448	0.060
A41	0.789	0.783	0.006
A42	0.572	0.546	0.026
A51	0.738	0.715	0.023
A52	0.632	0.595	0.037
D11	0.099	0.121	-0.022
D22	0.004	0.099	-0.095

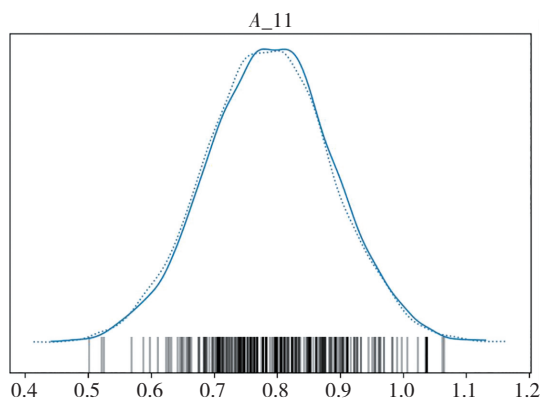
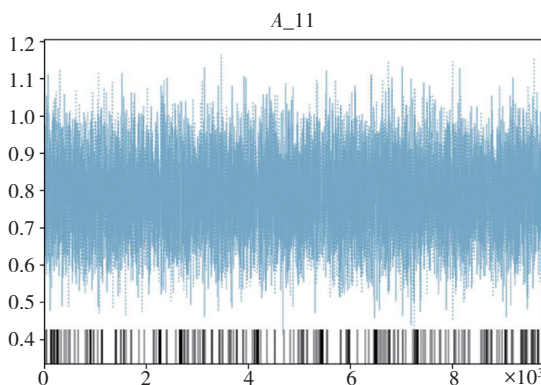
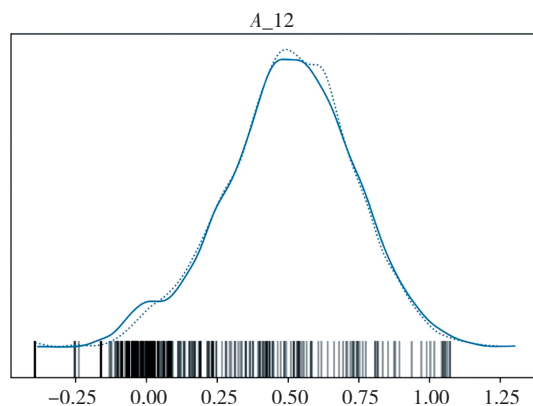
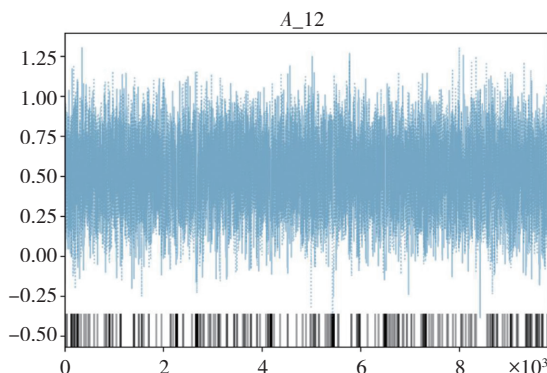
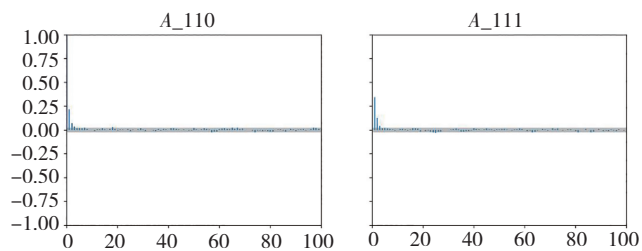
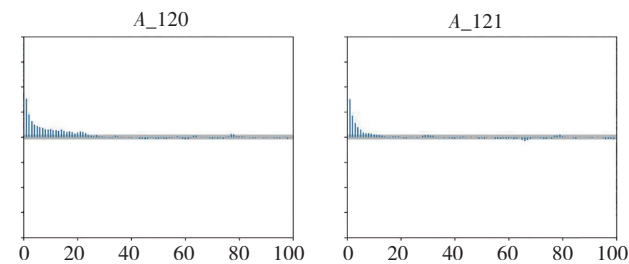
续表(表8)

参 数	主成分法参数值	VB 估计值	偏 差
D33	0.045	0.112	-0.068
D44	0.050	0.089	-0.039
D55	0.056	0.135	-0.079

为了保证两种算法的收敛性,绘制参数追踪图和自相关图来判断 MCMC 方法的收敛性,绘制 ELBO 图来判断 VB 方法的收敛性。

图 4 给出了当样本量 $N=713$ 时,MCMC 方法的参数追踪图。从图 4(a)、图 4(c) 可以看到:图中的两条曲线不仅是平滑的曲线,而且都趋近于高斯分布,说明核密度估计图表现良好。从图 4(b)、图 4(d) 可以看到:所有得到的采样值都没有强周期性和(特别是)位于一定趋势的区域内,说明路径图表现良好。

图 5 给出了当样本量 $N=713$ 时,MCMC 方法的自相关图,显示了采样值与相邻连续点(最多 100 个)之间的平均相关性,目的是判断 MCMC 方法的收敛性。可以看到所有生成的采样值之间没有很强的相关性,说明自相关图表现良好。

(a) 第一条链 A₁₁ 核密度估计图(b) 第一条链 A₁₁ 路径图(c) 第二条链 A₁₂ 核密度估计图(d) 第二条链 A₁₂ 路径图图 4 MCMC 方法的参数追踪图($N=713$)Fig. 4 Parameter trace plot of MCMC method ($N=713$)(a) 第一条链 A₁₁ 自相关图 (b) 第二条链 A₁₁ 自相关图(c) 第一条链 A₁₂ 自相关图 (d) 第二条链 A₁₂ 自相关图图 5 MCMC 方法的自相关图($N=713$)Fig. 5 Autocorrelation plot of MCMC method ($N=713$)

结合上述两种图形,可以判断 MCMC 方法是收敛的,这就保证了所有得到的采样值来自于正确的目标分布。

图 6 给出了当样本量 $N=713$ 时,VB 方法的 ELBO 图,目的是检验 VB 方法的收敛性。可以看到:随着迭代次数不断增加,变分下界 ELBO 在不断地收敛,于是认为 VB 方法是收敛的。

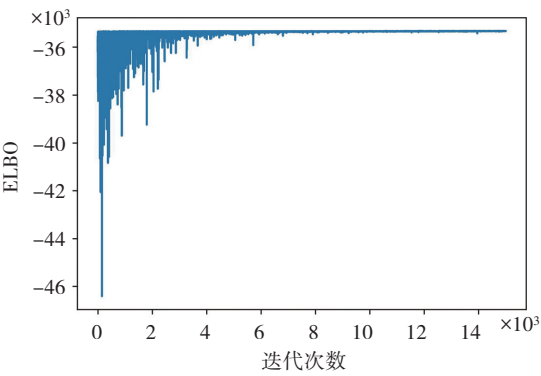


图 6 VB 方法的 ELBO 图($N=713$)

Fig. 6 ELBO plot of VB method ($N=713$)

由表 9 可知:当 $N=713$ 时,MCMC 方法和 VB 方法的运行时间和模拟结果一样,VB 方法运行的时间要比 MCMC 方法要短,当样本量不断增大时,这一时间差距会更加明显。因此 VB 方法的优势在于能节省更多时间。

表 9 MCMC 方法和 VB 方法的运行时间($N=713$)

Table 9 Running times of MCMC method and VB method ($N=713$)

样本量 N	MCMC 方法的	VB 方法的运行
	运行时间/s	时间/s
713	356	48.55

6 结论与展望

首先介绍了因子分析模型的提出和发展历程,通过已有的文献研究,发现因子分析模型广泛应用于贝叶斯统计、信息科学、行为科学、社会科学等领域中,而实际生活中的数据有时是非负的,可见,因子服从指数分布是一个合理的选择。

然后利用变分贝叶斯(VB)方法,对因子服从指数分布的因子分析模型进行了模型参数的估计。其中 VB 方法使用坐标上升变分推断(CAVI)算法对参数进行迭代求解,并与 MCMC 方法进行比较,通过随机模拟揭示了当样本量为 300 时,VB 方法和 MCMC 方法的有效性,绘制 ELBO 图判断了 VB 方法的收敛性,绘制参数追踪图、自相关图判断了 MCMC 方法的收敛性,并根

据预测值和真实值之间的偏差判断两种方法的好坏。通过随机模拟可以看到:两种方法都是收敛的,偏差的绝对值均小于 0.1,但是 VB 方法在估计精度、计算复杂度和运行时间上均优于 MCMC 方法,特别是当样本容量很大时,VB 方法的优势更加明显。

最后对样本量为 713 时的实际数据集进行了实证分析。结果表明:与 MCMC 方法相比,变分贝叶斯方法可以有效减少因子分析模型后验分布函数的计算量,估计参数的效果更好。VB 方法的 3 个优势:由于 VB 方法没有涉及复杂后验分布积分的计算,因此,计算更加简便;由于 VB 方法基于一个近似分布,当样本量很大时,其运行时间远小于 MCMC 方法;从运算结果来看,VB 方法的估计精度优于 MCMC 方法。

针对 VB 方法,接下来可以考虑在本文的模型及算法基础上,探究数据的缺失和高维以及变量选择问题,从而进一步推广因子分析模型的应用。

参考文献(References):

[1] WEI X, LI C. Bayesian mixtures of common factor analyzers: Model, variational inference, and applications [J]. Signal Processing, 2013, 93(11): 2894–2905.

[2] SHETH J N. Using factor analysis to estimate parameters [J]. Journal of the American Statistical Association, 1969, 64(327): 808–822.

[3] BAI J S, LI K P. Maximum likelihood estimation and inference for approximate factor models of high dimension [J]. Review of Economics and Statistics, 2016, 98(2): 298–309.

[4] 李顺静. 基于不完全数据的最大似然估计方法——EM 算法 [J]. 重庆工商大学学报(自然科学版), 2014, 31(5): 29–33.

LI Shun-jing. Maximum likelihood estimation method based on incomplete data: EM algorithm [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2014, 31(5): 29–33.

[5] GUO Y, CHANG W, ZHOU S. The research and simulation about rare events based on MCMC [C]//The 27th Chinese Control and Decision Conference (2015 CCDC). IEEE, 2015: 694–698.

[6] 岳博, 焦李成. 两阶段混合因子分析算法 [J]. 自然科学进展, 2002, 12(2): 219–222.

YUE Bo, JIAO Li-cheng. Two-stage mixed factor analysis algorithm [J]. Progress in Natural Science, 2002, 12(2): 219–222.

[7] 朱慧明, 孙雄志. 基于混合先验分布的贝叶斯因子分析模

- 型[J]. 湖南大学学报(自然科学版), 2007, 34(9): 82–85.
- ZHU Hui-ming, SUN Xiong-zhi. Bayesian factor analysis model based on mixed prior distribution[J]. Journal of Hunan University(Natural Science Edition), 2007, 34(9): 82–85.
- [8] PAPASTAMOULIS P. Overfitting bayesian mixtures of factor analyzers with an unknown number of components [J]. Computational Statistics & Data Analysis, 2018, 124(3): 220–234.
- [9] 夏业茂, 凌耀斌, 熊双桢. 两部分因子分析模型的贝叶斯推断[J]. 应用数学, 2018, 31(4): 761–778.
- XIA Ye-mao, LING Yao-bin, XIONG Shuang-can. Bayesian inference of a two-part factor analysis model [J]. Applied Mathematics, 2018, 31(4): 761–778.
- [10] ATTIAS H. A variational Bayesian framework for graphical models [J]. Advances in Neural Information Processing Systems, 1999, 12(1): 209–215.
- [11] 廖雪丽, 陈金叶, 张琪, 等. 两部分潜变量模型的变分贝叶斯推断[J]. 系统科学与数学, 2023, 43(4): 1039–1068.
- LIAO Xue-li, CHEN Jin-ye, ZHANG Qi, et al. Variational bayesian inference for two-part latent variable models [J]. System Science and Mathematics, 2023, 43(4): 1039–1068.
- [12] 徐定杰, 沈忱, 沈锋. 混合高斯分布的变分贝叶斯学习参数估计[J]. 上海交通大学学报(自然科学版), 2013, 47(7): 1119–1125.
- XU Ding-jie, SHEN Chen, SHEN Feng. Estimation of variable-decibel Bayesian learning parameters with mixed Gaussian distribution[J]. Journal of Shanghai Jiao Tong University(Natural Science Edition), 2013, 47(7): 1119–1125.
- [13] GIORDANO R, BRODERICK T, JORDAN M I. Covariances, robustness and variational bayes[J]. Journal of Machine Learning Research, 2018, 19(51): 1–59.
- [14] TZIKAS D G, LIKAS A C, GALATSANOS N P. The variational approximation for bayesian inference [J]. IEEE Signal Processing Magazine, 2008, 25(6): 131–146.
- [15] EDSON V, GEMANO D, Nunes A O, et al. Qeon questionnaire for assessing experiences in virtual learning environments[J]. IEEE Latin America Transactions, 2017, 15(6): 1197–1204.
- [16] 陈钊, 范剑青, 王丹. 高维因子模型及其在统计机器学习中的应用[J]. 中国科学: 数学, 2020, 50(4): 447–490.
- CHEN Zhao, FAN Jian-qing, WANG Dan. High dimensional factor model and its application in statistical machine learning [J]. Chinese Science: Mathematics, 2020, 50(4): 447–490.
- [17] ONATSKI A. Determining the number of factors from empirical distribution of eigenvalues [J]. The Review of Economics and Statistics, 2010, 92(4): 1004–1016.
- [18] 张璇. 分层线性模型的最大后验估计[J]. 统计与信息论坛, 2011, 26(1): 10–15.
- ZHANG Xuan. Maximum a posteriori estimation of hierarchical linear models [J]. Statistics and Information Forum, 2011, 26(1): 10–15.
- [19] YUN J, SHIN M, HOON JIN I, et al. Stochastic approximation hamiltonian monte carlo [J]. Journal of Statistical Computation and Simulation, 2020, 90(17): 3135–3156.
- [20] HOFFMAN M D, GELMAN A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo [J]. Journal of Machine Learning Research, 2014, 15(1): 1593–1623.
- [21] 周国琼, 蒋文江. 因子服从指数分布的因子分析模型的参数估计研究[J]. 云南师范大学学报(自然科学版), 2020, 40(6): 29–34.
- ZHOU Guo-qiong, JIANG Wen-jiang. Parameter estimation of factor analysis model with factor obeying exponential distribution[J]. Journal of Yunnan Normal University(Natural Science Edition), 2020, 40(6): 29–34.
- [22] 巴丽伟, 童常青. 基于变分贝叶斯推断的因子分析法[J]. 杭州电子科技大学学报(自然科学版), 2022, 42(3): 95–102.
- BA Li-wei, TONG Chang-qing. Factor analysis based on variational Bayesian inference[J]. Journal of Hangzhou Dianzi University(Natural Science Edition), 2022, 42(3): 95–102.
- [23] BLEI D M, KUCUKELBIR A, MCAULIFFE J D. Variational inference: A review for statisticians [J]. Journal of the American Statistical Association, 2017, 112(518): 859–877.
- [24] YOU C, ORMEROD J T, MUELLER S. On variational bayes estimation and variational information criteria for linear regression models[J]. Journal of Statistics, 2014, 56(1): 73–87.
- [25] WANG B, TITTERINGTON D M. Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model [J]. Bayesian Analysis, 2006, 1(1): 625–649.

责任编辑: 李翠薇