arXiv:2303.00473v1 [stat.ME] 1 Mar 2023

广义累积收缩过程先验及其在稀疏贝叶斯因子分析中的应用

Sylvia Fr¨uhwirth-Schnatter*

2023年3月2日

摘要

本文讨论了对一系列参数施加递增收缩的收缩先验。我们回顾了 Legramanti 等（2020）提出的累积收缩过程（CUSP）先验。这是一种尖峰与厚尾收缩先验，其尖峰概率随机递增，并由狄利克雷过程先验的打破棍表示构造而成。本文的第一项贡献是将这一 CUSP 先验扩展至涉及由贝塔分布产生的任意打破棍表示。第二项贡献是证明，可交换尖峰与连续分量收缩先验（在稀疏贝叶斯因子分析中颇受欢迎且应用广泛）可表示为有限广义 CUSP先验，而该先验可通过 slab 概率的降序次序统计量轻松获得。因此，随着载荷矩阵列索引增加，可交换尖峰与连续分量收缩先验会带来递增收缩，而无需对 slab 概率施加显式顺序约束。本文通过一个稀疏贝叶斯因子分析应用展示了这些研究发现的实用价值。本文还提出了一种基于 Cadonna 等人（2020）三重伽马先验的新型可交换尖峰与连续分量收缩先验，并通过模拟研究证明，该先验有助于估计未知的因子数量。

1 引言

收缩先验在现代贝叶斯推断中不可或缺，并使人们能够以原则性的方式处理模型设定不确定性。一个尤其重要的应用领域是贝叶斯因子分析，其中包含种类繁多且可能具有实用价值的收缩先验。

在因子分析中，假设维度为 m 的多元观测 $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})^\top, t = 1, \ldots, n$ 的协方差矩阵 $\mathbf{\Omega}_0 = \boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top + \mathbf{\Sigma}_0$ 由高斯因子

*金融、会计与统计系，维也纳经济与商业大学，奥地利。电子邮箱：sfruehwi@wu.ac.at。本文是主题专刊“贝叶斯推断：挑战、观点与前景”15篇投稿中的一篇，该专刊即将发表于《皇家学会哲学汇刊 *A* 辑》，DOI: 10.1098/rsta.2022.0148。

模型

$$\mathbf{y}_t = \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

其中，$\boldsymbol{\epsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_0)$ 是特异误差，且满足 $\boldsymbol{\Sigma}_0 = \text{Diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right)$；$\mathbf{f}_t \sim \mathcal{N}_{H_0}(\mathbf{0}, \mathbf{I})$ 是因子维度为 H_0的潜在因子。

在应用因子分析中，因子空间的维度 H_0 通常未知，必须根据数据推断。贝叶斯方法为解决这一问题提供了一种颇具吸引力的方案，因为未知因子维度 H_0 可以在过拟合因子模型中，与其他所有未知参数一并估计，例如载荷矩阵 $\boldsymbol{\beta}_0 \in \mathbb{R}^{m \times H_0}$ 中的因子载荷 $\{\beta_{ih}\}$ 以及特异方差 $\sigma_1^2, \ldots, \sigma_m^2$。在稀疏贝叶斯因子分析中，恢复因子数量的策略是：在因子模型的载荷矩阵中诱导出零列；该模型假定有$H > H_0$ 列，刻意使其过拟合真实但未知的因子维度 H_0。在有限因子分析中，选择 $H \leq (m-1)/2$ 以确保计量经济学识别（例如参见 Fr¨uhwirth-Schnatter 等人（2022a）以及 Hosszejni 和 Fr¨uhwirth-Schnatter（2022））；而在无限因子分析中，$H = \infty$；该领域的开创性工作可参见 Bhattacharya 和 Dunson（2011）。随后对因子载荷施加收缩先验，目的是根据数据中的信息自动移除所有冗余列。在稀疏贝叶斯因子分析中，选择收缩先验主要有四种方法。

一类文献在有限因子分析中采用连续收缩先验，通常着眼于协方差矩阵$\boldsymbol{\Omega}_0$, 的高效估计，参见例如 Kastner（2019）。虽然这些先验会隐式地缩减参数空间的维度，但如何明确地确定未知因子维度 H_0并不简单。

在无限因子分析中，Bhattacharya 和 Dunson（2011）同样对因子载荷采用连续收缩先验。他们提出了乘法伽马过程（MGP）先验，旨在惩罚因子载荷矩阵中额外列的影响。MGP先验将特定列中所有因子载荷的先验精度定义为伽马先验的累积乘积。该先验已得到广泛应用，例如 Murphy 等（2020）将其用于因子分析器的无限混合，De Vito 等（2021）则将其用于高通量生物数据的贝叶斯多研究因子分析。然而，Durante（2017）表明，只有在特定的超参数设置下，才能实现预期的递增收缩目标。因此，该方法往往会高估真实因子数，Legramanti 等（2020）通过一项全面的模拟研究对此进行了证实。

第三类极为丰富的文献采用可交换尖峰-板块先验，并为尖峰和厚尾分别设置列特异概率，参见 Fr¨uhwirth-Schnatter 等（2022b）；West（2003）；Carvalho 等（2008）；Teh 等（2007）；Fr¨uhwirth-Schnatter 和 Lopes（2010）；Conti 等（2014）；Ro˘ckov´a 和 George（2017）；Kaufmann 和 Schuhmacher（2019）等。具体而言，为载荷矩阵中的每个元素 β_{ih} 引入一个二元指示变量 δ_{ih}，并为每一列设定一个列特异的发生概率.....

假设因子载荷矩阵 h 每列中非零元素的发生概率 $\mathrm{P}(\delta_{ih}=1|\tau_h)=\tau_h$ 各不相同。与 MGP先验不同，所有列的厚尾概率$\tau_1,\ldots,\tau_H$采用可交换先验，且不对载荷矩阵的列施加明确的先验排序或递增收缩。有限贝叶斯因子分析中，此类可交换先验的一个常见例子是单参数贝塔先验 $\tau_h|H\sim\mathcal{B}\left(\frac{\alpha}{H},1\right)$ Ro˘ckov´a 和 George（2017）；Fr¨uhwirth-Schnatter 等（2022b）；Avalos-Pacheco等（2022）。与连续收缩先验不同，尖峰与 slab 先验的离散特性使得人们能够针对未知因子维度 H_0进行明确推断。

最后，作为对上述任一先验的替代，Legramanti 等（2020）近期提出了累积收缩过程（CUSP）先验。在无限因子模型中，CUSP先验是一种尖峰与厚尾先验，其中载荷矩阵的各列按顺序排列，且随着列索引增加，分配给尖峰的先验概率也随之增大。CUSP先验旨在体现这样一种预期：载荷矩阵中新增的列所发挥的作用会逐渐减弱，其相关参数的影响也呈随机递减趋势。构建CUSP先验时，Legramanti 等（2020）利用了狄利克雷过程（DP）先验的打破棍表示（Sethuraman，1994）。近期，Kowal 和 Canale（2022）从两个方面扩展了CUSP先验：首先，考虑更一般的尖峰分布；其次，采用 Teh 等（2007）提出的双参数印度自助餐过程先验的打破棍表示。作者在半参数函数因子模型中应用了这种“有序尖峰与厚尾先验”。

本文对这一研究领域作出了两项主要贡献。首先，本文将 Legramanti 等人（2020）和 Kowal 和 Canale（2022）的累积收缩过程先验推广至广义累积收缩过程先验这一类别，其中涉及非常一般的棍棒折断表示，这些表示可以是有限的，也可以是无限的。本文证明，对尖峰概率进行排序会使感兴趣参数的收缩程度递增；这一点已由 Kowal 和 Canale（2022，命题 1）针对有序尖峰与厚尾先验证明（该先验将 Legramanti 等（2020）的 CUSP 先验作为特例）。广义 CUSP 先验涵盖了文献中较早为因子分析模型提出的若干特定先验，这些先验涉及基于贝塔分布的棍棒折断表示，参见 Ro˘ckov´a and George（2017）；Heaukulani and Roy（2020）；Ohn and Kim（2022）；Fr¨uhwirth-Schnatter 等人（2022b）；Kowal 和 Canale（2022）。

其次，我们为广受欢迎的可交换收缩过程先验这一类先验带来了新的认识，其中包括一参数和二参数贝塔先验。我们证明，对一列参数施加的任意可交换尖峰与厚尾先验，都可以表示为广义累积收缩过程先验，并且会隐式地对这些参数施加递增收缩。这种表示可以通过 slab 概率的降序次序统计量简单推导得出。最后，我们讨论了这一广义 CUSP 先验在有限稀疏贝叶斯因子模型中的应用。

其余部　　本文的组织结构如下。第 2 节介绍广义累积收缩　　-

过程先验，并给出若干示例。第 3 节展示如何将可交换收缩过程先验表示为广义 CUSP 先验。第 4 节讨论这两类先验的后验推断。第 5 节说明其在稀疏贝叶斯因子分析中的应用，第 6 节进行总结。

2 广义累积收缩过程先验

2.1 定义

累积收缩过程（CUSP）先验由 Legramanti 等（2020）提出，用于对可数模型参数序列 $\{\theta_h\}$、$h = 1...H$施加递增收缩。递增收缩通过为每个参数 θ_h指定尖峰与厚尾先验来实现，

$$\theta_h|\pi_h \sim \pi_h\delta_{\{\theta_\infty\}} + (1-\pi_h)P_{\text{slab}}(\theta_h), \tag{2}$$

其中，将递增的先验概率 π_h 分配给位于 θ_∞处的狄拉克尖峰。基于狄利克雷过程（DP）的打破棍表示 ν_h $i.i.d.$ $\mathcal{B}$(1、 α)（Sethuraman，1994），递增的尖峰概率序列 π_h 定义如下：

$$\pi_h = \sum_{\ell=1}^{h}\omega_\ell, \quad \omega_\ell = \nu_\ell\prod_{j=1}^{\ell-1}(1-\nu_j). \tag{3}$$

显然，式（3）中的尖峰概率是递增的，因为 $\pi_h = \pi_{h-1} + \omega_h$，其中 $\omega_h \in$ （0、1）。对于 $H = \infty$，式（3）中定义的序列 $\{\omega_h\}$ 是狄利克雷过程混合的权重 $\{\omega_h\}$ 的打破棍表示。因此，$\sum_{\ell=1}^{\infty}\omega_\ell = 1$，并且随着 h 增加， π_h 趋近于 1。通过对某个有限的 $H < \infty$定义 $\nu_H = 1$，可以得到 CUSP先验的有限（截断）版本。

最近，Kowal 和 Canale（2022）提出了有序尖峰与厚尾先验，将式（2）中定义的 CUSP先验从多个方向进行了推广。首先，作者考虑了一般的尖峰分布，

$$\theta_h|\pi_h \sim \pi_h P_{\text{spike}}(\theta_h) + (1-\pi_h)P_{\text{slab}}(\theta_h), \tag{4}$$

而 Legramanti 等（2020）假设在一个已知的小值 $\delta_{\{\theta_\infty\}}$ 处存在狄拉克尖峰。然而，正如 Schiavon和Canale（2020）在无限因子模型背景下所示，该参数的选择可能会产生非常显著的影响。

其次，Kowal 和 Canale（2022）以更一般的方式构造了式（3）中的递增尖峰概率序列 $\{\pi_h\}$， h =1，.....， ∞，即将其构造为一个累积过程，该过程涉及 Teh 等（2007）提出的双参数印度自助餐过程（IBP）先验的打破棍表示 ν_h $i.i.d.$ $\mathcal{B}$(β， $\beta\alpha$)。当 $\beta = 1$时，得到打破棍表示 ν_h $i.i.d.$ $\mathcal{B}$(1， α)，有序尖峰与厚尾先验便简化为 Legramanti 等人（2020）的 CUSP 先验。

CUSP 先验中的强度参数 α 在决定有多少参数 θ_h 处于激活状态方面起着重要作用；Legramanti 等人（2020）假设该参数已知且固定为 $\alpha = 5$。与此不同，Kowal 和 Canale（2022）允许将 α（在其论文中称为 κ）视为未知超参数，并通过伽马先验 $\alpha \sim \mathcal{G}(a^\alpha,\ b^\alpha)$，根据数据进行学习；而 β（在其论文中称为 ι）则固定不变，通常取为 $\beta = 1$。

本文进一步拓展了这项重要工作，并在定义 2.1 中引入了广义 CUSP 先验。具体而言，递增的尖峰概率序列 $\{\pi_h\},\ h = 1,\ 2,\ \ldots\ldots$ 是通过累积过程构造的，该过程涉及更一般的（也可能是有限的）断棒构造 $\{\nu_h\}$，这些构造不必来自同一分布。Legramanti 等（2020）以及 Kowal 和 Canale（2022）提出的 CUSP 先验是这种广义 CUSP 先验的特例。第 2.2 节和第 3.2 节还讨论了其他示例。

定义 2.1（广义累积收缩过程先验）。设 $\boldsymbol{\pi} = \{\pi_h \in (0,\ 1)\}, h = 1, \ldots, H$ 是一个取值于单位区间的可数随机参数序列，其定义如下：

$$\pi_h = \sum_{\ell=1}^{h} \omega_\ell, \quad \omega_\ell = \nu_\ell \prod_{j=1}^{\ell-1}(1-\nu_j), \tag{5}$$

其中，$\{\nu_h\}, h = 1, \ldots, H$ 是一个取值于单位区间的随机变量序列。对于 $H < \infty, \nu_H$，可以取值 *1*，但不一定必须取值 *1*。对于 $H = \infty$，假设 $\sum_{\ell=1}^{\infty} \omega_\ell = 1$ 几乎必然成立。设 $\Theta = \{\theta_h\}, h = 1, \ldots, H$ 为一个可数的模型参数序列。假设给定 $\boldsymbol{\pi}$ 时，参数 θ_h 相互独立，并且对于所有 h，$\theta_h|\pi_h$ 与 $\pi_\ell, \ell \neq h$ 相互独立。假设 $p(\theta_h|\pi_h)$ 具有以下尖峰与厚尾先验的形式：

$$\theta_h|\pi_h \sim \pi_h P_{\text{spike}}(\theta_h) + (1-\pi_h) P_{\text{slab}}(\theta_h). \tag{6}$$

于是，对于 $H < \infty$，Θ 被称为服从有限广义累积收缩过程（*CUSP*）先验。如果 $H = \infty$，则 Θ 被称为服从无限广义 *CUSP* 先验。

注意，定义（5）中的尖峰概率 $\pi_h = \pi_{h-1} + \omega_h$ 按照构造是一个递增序列，并且 $\mathbb{E}(\pi_h) > \mathbb{E}(\pi_{h-1})$。广义CUSP先验中尖峰概率 $\{\pi_h\}$ 的排序意味着参数 $\{\theta_h\}$ 在（6）中的先验分布也具有明确的排序。这一点已由（Kowal 和 Canale，2022，命题 1）针对有序尖峰与厚尾先验加以证明，并扩展了（Legramanti等人，2020，引理1）的结果。通过对相应证明进行直接扩展可知，这一重要性质同样适用于定义 2.1 中引入的广义 CUSP 先验。这一见解概括于命题 2.1 中。

命题 2.1。对于 $\varepsilon > 0$ 和固定的 θ_0，令 $\mathbb{B}_\varepsilon(\theta_0) = \{\theta_h :\ |\theta_h - \theta_0| < \varepsilon\}$。在先验（*5*）下，只要（*6*）中的尖峰和厚尾分布满足

$$P_{\text{spike}}(\mathbb{B}_\varepsilon(\theta_0)) > P_{\text{slab}}(\mathbb{B}_\varepsilon(\theta_0)), \tag{7}$$

then

$$\mathrm{P}(|\theta_h - \theta_0| \le \varepsilon) < \mathrm{P}(|\theta_{h+1} - \theta_0| \le \varepsilon). \tag{8}$$

证明是将（Kowal 和 Canale，2022，命题 1）的相应证明直接推广到广义 CUSP 先验。对于任意 $\varepsilon > 0$ 和固定的 θ_0，以下结论成立：

$$\begin{aligned}\mathrm{P}(|\theta_h - \theta_0| \le \varepsilon) &= P_{\text{slab}}(\mathbb{B}_\varepsilon(\theta_0))\mathbb{E}(1-\pi_h) + P_{\text{spike}}(\mathbb{B}_\varepsilon(\theta_0))\mathbb{E}(\pi_h) \\ &= P_{\text{spike}}(\mathbb{B}_\varepsilon(\theta_0)) + (1-\mathbb{E}(\pi_h))\left(P_{\text{slab}}(\mathbb{B}_\varepsilon(\theta_0)) - P_{\text{spike}}(\mathbb{B}_\varepsilon(\theta_0))\right).\end{aligned}$$

Sin由于 $\mathbb{E}(\pi_h)$ 关于 h 严格递增，（若 (7) 成立，则）可立即得到 (8)：

$$\mathrm{P}(|\theta_{h+1} - \theta_0| \le \varepsilon) - \mathrm{P}(|\theta_h - \theta_0| \le \varepsilon) = (\mathbb{E}(\pi_{h+1}) - \mathbb{E}(\pi_h))\left(P_{\text{spike}}(\mathbb{B}_\varepsilon(\theta_0)) - P_{\text{slab}}(\mathbb{B}_\varepsilon(\theta_0))\right) > 0.$$

验证 slab 概率的降序序列 $\pi_h^\star = 1 - \pi_h$ 具有如下表示也很有意思：

$$\pi_h^\star = \prod_{\ell=1}^{h}(1-\nu_\ell) = \prod_{\ell=1}^{h}\nu_\ell^\star, \qquad h = 1, 2, \ldots, H. \tag{9}$$

利用归纳法很容易证明这一结果。显然，(9) 对于 $h = 1$ 成立，因为 $\pi_1^\star = 1 - \pi_1 = 1 - \omega_1 = 1 - \nu_1 = \nu_1^\star$。假设 (9) 对直到 $h-1$ 的情形均成立。那么 $\pi_h = \pi_{h-1} + \omega_h$，其中

$$\omega_h = \nu_h \prod_{\ell=1}^{h-1}(1-\nu_\ell) = \nu_h \pi_{h-1}^\star,$$

于是得到：

$$\pi_h^\star = 1 - \pi_h = 1 - \pi_{h-1} - \omega_h = \pi_{h-1}^\star - \nu_h\pi_{h-1}^\star = (1-\nu_h)\pi_{h-1}^\star = \prod_{\ell=1}^{h}\nu_\ell^\star.$$

2.2 CUSP 先验示例

定义 2.1 相当一般，并未对随机变量序列 $\{\nu_h\}, h = 1, \ldots\ldots, H$ 作出任何具体假设。在第 3 节中，我们将展示如何从有限可交换收缩过程先验中 slab 概率的降序次序统计量推导出 $\{\nu_h\}$。

或者，也可以选择使折棒 $\{\nu_h\}$ 来自特定的分布族。例如，它们可以作为相互独立的随机变量，分别服从贝塔分布：

$$\nu_h \sim \mathcal{B}(a_h, b_h), \quad h = 1, \ldots, H. \tag{10}$$

利用（9），递减的厚尾概率 $\pi_h^\star$ 可表示为具有$\nu_h^\star \sim \mathcal{B}$（$b_h$，$a_h$）的乘法贝塔过程。

文献中已经提出了此类收缩先验的几种特殊情形。显然，Legramanti 等（2020）引入的CUSP先验是（10）在 $H = \infty$、$a_h = 1$和 $b_h = \alpha$时得到的特殊情形。正如Teh 等（2007）所指出的，该先验等价于印度自助餐过程先验。Ghahramani等人（2007）从打破棍表示 $\nu_h \; i.i.d. \; \mathcal{B}$（$\beta$，$\beta\alpha$）定义了双参数印度自助餐过程先验，这是对Teh 等（2007）印度自助餐过程先验的扩展。Kowal 和 Canale（2022）的有序尖峰与厚尾先验基于这一打破棍表示，并且在 $H = \infty$、$a_h = \beta$ 和 $b_h = \beta\alpha$时成为（10）的特殊情形。在高维稀疏因子模型的背景下，Ohn 和 Kim（2022）按照（9）在因子载荷的尖峰与厚尾先验中定义了厚尾概率序列 $\pi_h^\star$，其中 $\nu_h^\star \sim \mathcal{B}$（$\alpha$，$1+\kappa$），$\alpha > 0$ 和 $\kappa \geq 0$ 为超参数。该先验在 $H = \infty$、$a_h = 1 + \kappa$ 和 $b_h = \alpha$时成为（10）的特殊情形。此外，它还导出了一种双参数印度自助餐过程先验，其参数化方式与Kowal 和 Canale（2022）所采用的先验存在替代参数化。

构造广义 CUSP 先验的另一种方法，是利用比狄利克雷过程混合更一般的混合分布的权重。原则上，任何有限或无限混合分布的权重都可用于定义广义 CUSP 先验，参见（Teh 等人，2007，图 1）。例如，Pitman-Yor过程（PYP）先验已被 Heaukulani 和 Roy（2020）用于定义吉布斯型印度自助餐过程。选择 $\nu_h \sim \mathcal{B}(1-\sigma,\ \alpha + h\sigma)$，其中 $\sigma \in [0,1)$ 和 $\alpha > \sigma$ 在（10）中，意味着会有大量活跃系数，其权重 π_h^*虽显著但较小。对于 $\sigma = 0$，由此诱导的广义 CUSP 先验会简化为 Legramanti 等人（2020）的 CUSP 先验。或者，也可以选择 PYP 先验，即 $\nu_h \sim \mathcal{B}(1-\sigma,\ (H-h)|\sigma|)$、$h = 1, \ldots\ldots, H-1$，其中 $\sigma < 0$ 为负数，H 为自然数。在（10）中采用这种打破棍表示，会得到一种广义 CUSP，其中活跃系数的数量 H 有限，且 $a_h = 1 - \sigma$ 和 $b_h = (H-h)|\sigma|$。

3 可交换收缩过程先验

3.1 定义

定义 3.1 (可交换收缩过程先验)。设 $\boldsymbol{\tau} = \{\tau_h \in (0,\ 1)\}, h = 1, \ldots, H \; with H < \infty$ 是一个有限的独立同分布随机参数序列，其取值位于单位区间内。设 $\Theta = \{\theta_h\}, h = 1, \ldots, H$ 为一个有限的模型参数序列，并假设参数 $\theta_h|\tau_h$ 在给定 $\boldsymbol{\tau}$ 的条件下相互独立，且对所有 h 均独立于 τ_ℓ，$\ell \neq h$。如果 $p(\theta_h|\tau_h)$ 具有尖峰与厚尾先验的形式：

$$\theta_h|\tau_h \sim (1-\tau_h)P_{\text{spike}}(\theta_h) + \tau_h P_{\text{slab}}(\theta_h), \tag{11}$$

那么 Θ 被称为服从可交换收缩过程（*ESP*）先验。

根据定义，先验（11）在对 θ_h 的索引进行排列时保持不变。因此，如果序列 $\{\theta_h\}, h =$

$1, \ldots, H$蕴含一个可交换收缩过程（ESP）先验，那么对于索引 $1, \ldots, H$ 的任意排列 $\rho(1) \ldots, \rho(H)$，序列 $\{\theta_{\rho(h)}\}, h = 1, \ldots, H$服从相同的 ESP 先验。

为了完整定义 ESP 先验，必须为厚尾概率 $\tau_1, \ldots, \tau_H$指定一个概率分布。通常假设

$$\tau_h | H \sim \mathcal{B}(a_0, b_0), \qquad h = 1, \ldots, H, \tag{12}$$

其中 a_0 和 b_0 可能取决于 H 以及未知的超参数。

表示为 CUSP 先验。本文的一项主要贡献是证明，任意 ESP 先验都具有一个有限的广义 CUSP 表示，如 (5) 和 (6) 所示。CUSP 表示只需对索引 $1, \ldots, H$进行简单排列即可得到。考虑先验 (11) 中无序厚尾概率 $\tau_1, \ldots, \tau_H$的降序次序统计量 $\tau_{(1)} > \cdots > \tau_{(H)}$。如果按照降序厚尾概率 $\tau_{(1)}, \ldots, \tau_{(H)}$对系数 $\theta_1, \ldots, \theta_H$进行排列，那么广义 CUSP 表示中的尖峰概率 π_h 对于 $h = 1, \ldots, H$ 等于 $\pi_h = 1 - \tau_{(h)}$，并且根据定义是递增的。因此，根据命题 2.1，ESP 先验会对有序系数序列 $\theta_{\rho(1)}, \ldots, \theta_{\rho(H)}$产生递增收缩，其中参数 $\theta_1, \ldots, \theta_H$ 按照形成降序次序统计量 $\tau_{(1)}, \ldots, \tau_{(H)}$的排列顺序排序。因此，无需在 ESP 先验的定义中显式规定尖峰概率的任何顺序，即可实现递增收缩。

需要强调的是，要实现这一收缩性质，我们并不需要知道显式的 CUSP 表示。理论上，我们可以根据降序次序统计量$\tau_{(h)} < \tau_{(h-1)}$，从式（9）推导出折棒 $\nu_h^\star$ 或等价地ν_h 的分布：

$$\nu_h^\star = \frac{\tau_{(h)}}{\tau_{(h-1)}}, \qquad \nu_h = 1 - \frac{\tau_{(h)}}{\tau_{(h-1)}}.$$

然而，只有在特定情况下，才能求出折棒$\{\nu_h\}, h = 1, \ldots, H$的显式分布，示例见第 3.2 节。无论如何，$\pi_h = 1 - \tau_{(h)}$ 是一个递增序列，因此 $\mathbb{E}(\pi_{h+1}) > \mathbb{E}(\pi_h)$。要证明有序系数序列 $\theta_{\rho(1)}, \ldots, \theta_{\rho(H)}$ 满足命题 2.1，这些就足够了。

3.2 ESP 先验示例

许多作者都应用了可交换收缩过程先验，尤其是在稀疏贝叶斯因子分析中。例如，Fr¨uhwirth-Schnatter等人（2022b）假设通过选择 $a_0 = \frac{\alpha}{H}\beta$ 和 $b_0 = \beta$，使式（12）中的超参数 a_0 依赖于 H：

$$\tau_h | H \sim \mathcal{B}\left(\frac{\alpha}{H}\beta, \beta\right), \tag{13}$$

其中， H 是潜在因子的最大数量， α 和 β 是可从数据中估计的超参数。对于 $H \to \infty$，有限双参数贝塔先验（2PB）（13）收敛于Ghahramani等人（2007）在贝叶斯非参数潜在特征模型背景下提出的无限2PB先验。这些模型可视为含有无限多列的因子模型，其中只有有限列非零。

对于 $\beta = 1$，Ro˘ckov´a 和 George（2017）采用的有限单参数贝塔先验（1PB）得到：

$$\tau_h | H \sim \mathcal{B}\left(\frac{\alpha}{H}, 1\right). \tag{14}$$

众所周知，对于 $H \to \infty$，该先验收敛于印度自助餐过程（IBP）先验，参见 Teh 等人（2007）。附录中证明，先验（14）的广义 CUSP 表示涉及打破棍表示 $\nu_h \sim \mathcal{B}\left(1, \alpha^{\frac{H-h+1}{H}}\right)$，因此它是广义 CUSP 先验（10）的一种特殊情形。因此，当 H 趋于无穷时，有限1PB先验（14）收敛于 Legramanti 等（2020）提出的强度参数为 α的 CUSP 先验。

4 后验推断

ESP 先验和广义 CUSP 先验的后验推断均基于马尔可夫链蒙特卡罗（MCMC）估计，其中数据增强尤其有用。

根据应用场景的不同， θ_h 通常充当层次先验的超参数$\boldsymbol{\beta}_h|\theta_h$，该层次先验 $\boldsymbol{\beta}_h$ 涉及额外的模型参数 $\boldsymbol{\beta}_h$（例如列特异性因子载荷 $\boldsymbol{\beta}_h = (\beta_{1h}, \ldots\ldots, \beta_{mh})^\top$）。对 θ_h 进行边际化可得到以下针对 $\boldsymbol{\beta}_h$的尖峰与厚尾先验：

$$\boldsymbol{\beta}_h | \pi_h \sim \pi_h P_{\text{spike}}(\boldsymbol{\beta}_h) + (1 - \pi_h) P_{\text{slab}}(\boldsymbol{\beta}_h), \tag{15}$$

其中，尖峰和厚尾分布的概率密度函数分别为：

$$p_{\text{spike}}(\boldsymbol{\beta}_h) = \int p(\boldsymbol{\beta}_h|\theta_h) p_{\text{spike}}(\theta_h)\, d\theta_h, \quad p_{\text{slab}}(\boldsymbol{\beta}_h) = \int p(\boldsymbol{\beta}_h|\theta_h) p_{\text{slab}}(\theta_h)\, d\theta_h.$$

4.1 ESP先验的数据增广与马尔可夫链蒙特卡罗

关于可交换收缩过程（ESP）先验的数据增广和马尔可夫链蒙特卡罗已有大量研究。对于先验 (11)，针对每个 $h = 1, \ldots, H$，引入一个二元指示变量 S_h，其伯努利先验为 P($S_h = 1|\tau_h$) $= \tau_h$。给定 S_h，参数 θ_h 随后在先验上被归类为尖峰或厚尾：

$$\theta_h | S_h \sim (1 - S_h) P_{\text{spike}}(\theta_h) + S_h P_{\text{slab}}(\theta_h).$$

在 MCMC 方案中，指示变量 $S_1, \ldots, S_H$ 以及 slab 概率 $\tau_1, \ldots, \tau_H$ 均作为未知量引入，并从各自的条件后验分布中抽样。

对指示变量 S_h 的采样在 $H \times 2$ 网格上进行，可以通过多种方式实现。按照 Legramanti 等（2020）的思路，可以在给定参数 $\boldsymbol{\beta}_1$、……、$\boldsymbol{\beta}_H$ 和厚尾概率 τ_1、……、τ_H 的条件下进行分类，

$$\mathrm{P}(S_h = 0 | \boldsymbol{\beta}_h, \tau_h) \propto (1 - \tau_h) p_{\text{spike}}(\boldsymbol{\beta}_h), \quad \mathrm{P}(S_h = 1 | \boldsymbol{\beta}_h, \tau_h) \propto \tau_h p_{\text{slab}}(\boldsymbol{\beta}_h),$$

其中，$p_{\text{spike}}(\boldsymbol{\beta}_h)$ 和 $p_{\text{slab}}(\boldsymbol{\beta}_h)$ 分别是式（15）中尖峰和厚尾分布的概率密度函数。通过两项修改可以得到更高效的采样器。首先，对 S_h 进行边际化采样，边际化对象为 τ_1、……、τ_H。其次，不使用式（15）中的多元混合分布（随着 $\boldsymbol{\beta}_h$ 的维度增加，该分布会变得相当有信息量），而是利用 θ_h 上的混合先验（11），直接基于 θ_1、……、θ_H 进行分类。这些修改得到：

$$\mathrm{P}(S_h = 0 | \theta_h, q_A) \propto (1 - q_A) \cdot p_{\text{spike}}(\theta_h), \quad \mathrm{P}(S_h = 1 | \theta_h, q_A) \propto q_A \cdot p_{\text{slab}}(\theta_h), \tag{16}$$

其中，$q_A = \mathbb{E}(\tau_h) = \frac{a_0}{a_0 + b_0}$ 是厚尾的先验概率期望，而 $p_{\text{spike}}(\theta_h)$ 和 $p_{\text{slab}}(\theta_h)$ 分别是式（11）中尖峰和厚尾分布的概率密度函数。

无论如何，给定指示变量 $S_1, \ldots, S_H$ 后，通过从 $\tau_h | S_h$ 中抽样 τ_h（$h = 1, \ldots, H$），即可更新 slab 概率 $\tau_1, \ldots, \tau_H$。在先验（12）下，这得到

$$\tau_h | S_h \sim \mathcal{B}\left(a_0 + S_h, b_0 + 1 - S_h\right). \tag{17}$$

4.2 CUSP先验的数据增广与马尔可夫链蒙特卡罗

为对 CUSP 先验执行马尔可夫链蒙特卡罗，Legramanti 等（2020）将无限表示（3）在 $H < \infty$ 处截断，并引入 h 个分类指标变量 $z_1, \ldots, z_H$。每个 z_h 取值于 $\{1, 2, \ldots, H\}$，其离散先验分布为 $\mathrm{P}(z_h = \ell) = \omega_\ell$，$\ell = 1, \ldots, H$。给定 z_h 后，尖峰与厚尾先验（6）可表示为：

$$\theta_h | z_h \sim \mathbb{I}\{z_h \leq h\} P_{\text{spike}}(\theta_h) + (1 - \mathbb{I}\{z_h \leq h\}) P_{\text{slab}}(\theta_h). \tag{18}$$

这种数据增广技术具有通用性，可直接应用于定义 2.1 中引入的广义 CUSP 先验，无需任何修改。

此外，Legramanti 等（2020）将折棒 $\nu_1, \ldots, \nu_H$ 作为未知量引入，并根据分类指标变量 $z_1, \ldots, z_H$ 从各自的条件后验分布中进行抽样。这一步很容易扩展到由贝塔分布产生的棍棒折断表示 $\nu_\ell \sim \mathcal{B}(a_\ell, b_\ell)$ 所诱导的广义 CUSP 先验；另见 Kowal 和 Canale（2022）。折棒 $\nu_1, \ldots, \nu_H$ 的更新

在给定指示变量 $z_1,\ldots,z_H$的条件下，通过对 $\ell=1,\ldots,H$对应的 ν_ℓ从 $\nu_\ell|z_1,\ldots,z_H$中抽样：

$$\nu_\ell|z_1,\ldots,z_H \sim \mathcal{B}\left(a_\ell+\sum_{h=1}^{H}\mathbb{I}\{z_h=\ell\}, b_\ell+\sum_{h=1}^{H}\mathbb{I}\{z_h>\ell\}\right).$$

对于 $(a_\ell,b_\ell)=(1,\alpha)$ 和 $(a_\ell,b_\ell)=(\beta,\beta\alpha)$，抽样步骤分别如 (Legramanti等人，2020，算法 1) 和 (Kowal 和 Canale，2022，算法 1) 所示。给定折棒 $\nu_1,\ldots,\nu_H$，权重$\omega_1,\ldots,\omega_H$和尖峰概率 $\pi_1,\ldots,\pi_H$依据 (5) 进行更新。

分类指标变量 $z_1,\ldots,z_H$的抽样在 $H\times H$网格上进行，条件是给定参数 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_H$和权重 $\omega_1,\ldots,\omega_H$：

$$\mathrm{P}(z_h=\ell|\boldsymbol{\beta}_h)\propto\begin{cases}\omega_\ell\, p_{\mathrm{spike}}(\boldsymbol{\beta}_h), & \ell=1,\ldots,h,\\ \omega_\ell\, p_{\mathrm{slab}}(\boldsymbol{\beta}_h), & \ell=h+1,\ldots,H.\end{cases}$$

给定指示变量 $z_1,\ldots,z_H$，系数 $\theta_1,\ldots,\theta_H$分别从尖峰或厚尾分布中抽样，采用 (18) 中给出的表示 $\theta_h|z_h$。

CUSP先验还可以表示为涉及二元指示变量 S_1、.....、S_H的形式，这些变量定义为 $S_h=\mathbb{I}\{z_h>h\}$，其先验概率为 $\mathrm{P}(S_h=1|\pi_h)=\mathrm{P}(z_h>h|\pi_h)=1-\pi_h=\pi_h^*$，其中 π_h^*是厚尾概率。给定 S_h后，先验（18）可以改写为：

$$\theta_h|S_h\sim(1-S_h)P_{\mathrm{spike}}(\theta_h)+S_hP_{\mathrm{slab}}(\theta_h).$$

人们可能会认为，与 ESP 先验一样，CUSP 先验的马尔可夫链蒙特卡罗方案中的分类变量 z_1、.....、z_H可以由二元指示变量 S_1、.....、S_H替代。这样可以大大简化抽样过程。然而，尽管可以采用与第 4.1 节类似的方式对 S_1、.....、S_H进行抽样，但无法基于二元指示变量 S_1、.....、S_H对折棒进行抽样，因为先验 $p(S_h|\pi_h^*)$ 仅包含关于事件 $\{z_h>h\}$ 和 $\{z_h\leq h\}$的信息，而不包含关于$\{z_h=\ell\}$ 和 $\{z_h>\ell\}$ （对于 $\ell\neq h$）的信息。

不过，采用有限可交换收缩过程先验可以提升计算效率；尤其当马尔可夫链蒙特卡罗估计基于无限 CUSP 先验的截断版本时更是如此，示例见第 5.2 节。在这种情况下，任何收敛到无限 CUSP 先验的有限可交换收缩过程先验都可以采用。例如，有限1PB先验（14）收敛到 Legramanti 等（2020）的 CUSP 先验，而有限 2PB 先验（13）收敛到 Kowal 和 Canale（2022）的有序尖峰与厚尾先验。

4.3 活跃系数数量的推断

在系数序列上引入 ESP 或 CUSP 先验，主要原因之一是为了学习有多少个系数处于激活状态。Legramanti 等人（2020）使用这一方法来估计

CUSP 先验（2）下稀疏贝叶斯因子分析中的激活因子数量；Kowal 和 Canale（2022）则将其用于估计采用有序尖峰与厚尾先验（4）进行贝叶斯秩选择的贝叶斯半参数函数因子模型中的激活项数量。

分别基于分类变量 z_1、……、 z_H 或二元指示变量 S_1、……、 S_H，活跃系数的数量 $H^\star$ 定义为:

$$H^\star = \sum_{h=1}^{H} \mathbb{I}\{z_h > h\}, \qquad H^\star = \sum_{h=1}^{H} S_h. \tag{19}$$

表示（19）有助于研究超参数的选择如何影响 $H^\star$ 的先验分布。正如 Legramanti 等（2020）所示，在 CUSP先验下，强度参数 α 会显著影响模型维度 $H^\star$ 的先验，其中均值 $\mathbb{E}(H^\star|\alpha)$ 和方差 $\mathbb{V}(H^\star|\alpha)$ 均等于 α。因此，Kowal 和 Canale（2022）建议，应根据数据学习 CUSP先验的超参数。

类似地，在有限 ESP 先验下，只有通过适当选择超参数才能实现递增收缩。例如，对于具有 $H < \infty$ 的有限1PB先验（14）， $H^\star$ 的均值和方差都强烈依赖于 α:

$$\mathbb{E}(H^\star|\alpha) = \frac{\alpha}{1+\alpha/H}, \qquad \mathbb{V}(H^\star|\alpha) = \frac{\alpha}{(1+\alpha/H)^2}.$$

当我们考察基于厚尾概率降序次序统计量 $\tau_{(1)} > \cdots > \tau_{(H)}$ 的有限1PB先验的 CUSP 表示时，α 的影响会更加明显。最大的厚尾概率 $\tau_{(1)}$ 服从 $\mathcal{B}(\alpha, 1)$ 分布，而后续的厚尾概率 $\tau_{(h)} = \tau_{(h-1)}\nu_h$， $\nu_h \sim \mathcal{B}(\alpha((H-h+1)/H), 1)$，会随着 h 增大而越来越趋近于零，最后的折棒 $\nu_{H-1} \sim \mathcal{B}(2\alpha/H, 1)$ 和 $\nu_H \sim \mathcal{B}(\alpha/H, 1)$ 也如此。因此，具有 $\alpha << H$ 的先验会诱导先验稀疏性，因为最大的尖峰概率 $\eta_h = 1 - \tau_{(h)}$ 会越来越趋近于 1。

一种不会诱导先验稀疏性的常见先验是均匀先验 $\tau_h \sim \mathcal{U}[0,1], h = 1$、……、 H。例如，Zhao 等（2016）应用该先验引入了具有列特异的固定载荷的狄拉克尖峰，这与 Legramanti等人（2020）对 CUSP先验的做法相同。形式上，均匀先验可视为 1PB 先验的一种特殊情形，其中$\alpha = H$。在这种情况下，最后三个折棒分别服从 $\nu_{H-2} \sim \mathcal{B}(3, 1)$、 $\nu_{H-1} \sim \mathcal{B}(2, 1)$ 和$\nu_H \sim \mathcal{B}(1, 1)$，因此，最大的三个尖峰概率先验上并不会明显趋近于 1。因此，这种先验容易导致活跃系数数量过拟合，这一点将在第 5.2 节的说明性案例研究中得到验证。

对于有限 ESP 先验，通常假设超参数是已知的，但也可以很容易地将其视为未知参数，并根据数据进行学习，参见例如 Fr¨uhwirth-Schnatter 等（2022b）。对于这两类收缩先验，我们将在适当先验下如何在 MCMC 估计过程中对超参数进行抽样的问题上，于第 4.4 节中展开更详细的讨论。

表示式（19）还有助于根据给定数据 $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ 推导活跃系数数量的后验分布 $p(H^\star|\mathbf{y})$。由于分类变量 $z_1, \ldots, z_H$ 和二元指示变量 $S_1, \ldots, S_H$ 是在 MCMC 方案中进行抽样的，因此借助（19）可以立即得到活跃系数数量的后验分布 $p(H^\star|\mathbf{y})$ 的抽样结果。

4.4 学习超参数

对于有序尖峰与厚尾先验（4），其中 $\nu_h \; i.i.d. \; \mathcal{B}(\beta, \beta\alpha)$，Kowal 和 Canale（2022）对强度参数 α 设定伽马先验 $\alpha \sim \mathcal{G}(a^\alpha, b^\alpha)$，而另一个超参数固定为 $\beta = 1$，从而将该先验简化为强度参数未知的 CUSP先验。实践中，$a^\alpha = 2$ 和$b^\alpha = 1$ 的取值经过选择，使得 $\mathbb{E}(H^*) = \mathbb{V}(H^*) = 2$。为实现高效的 MCMC 估计，Kowal 和 Canale （2022）将 CUSP 先验截断于 $H < \infty$，并假设 $\nu_H = 1$。在这些假设下，伽马先验与折棒 $\nu_1, \ldots, \nu_{H-1}$ 的似然条件共轭，因此 $\alpha|v_1, \ldots, v_{H-1}$ 可以从伽马分布中轻松更新

$$\alpha|v_1, \ldots, v_{H-1} \sim \mathcal{G}(a^\alpha + H - 1, b^\alpha - \sum_{h=1}^{H-1} \log(1 - v_h)). \tag{20}$$

对于有限 ESP 先验，我们发现，使用边缘化后验分布（将 slab 概率 $\tau_1, \ldots, \tau_H$ 积分掉）来学习未知超参数更为合适。对于基于贝塔先验 (12) 的有限 ESP 先验，这很容易实现，其中 $\tau_h|H \sim \mathcal{B}(a_0, b_0)$。在该先验下，似然 $p(S_1, \ldots, S_H|a_0, b_0)$ 可写成闭式，且对指示变量 $S_1, \ldots, S_H$ 的依赖仅通过活跃系数的数量 $H^\star$ 来体现，该数量定义于 (19)：

$$p(S_1, \ldots, S_H|a_0, b_0) = (q_A)^{H^\star}(1 - q_A)^{H - H^\star}, \quad q_A = \mathrm{P}(S_h = 1) = \frac{a_0}{a_0 + b_0}.$$

只要 a_0 和/或 b_0 依赖于未知超参数（例如式 (14) 中定义的有限 1PB 先验里的 α），就可以将该似然与合适的先验结合起来。在这种情况下，$q_A = \alpha/(\alpha + H)$ 以及伽马先验 $\alpha \sim \mathcal{G}(a^\alpha, b^\alpha)$ 下的后验分布 $p(\alpha|S_1, \ldots, S_H) = p(\alpha|H^\star)$ 为：

$$p(\alpha|H^\star) \propto p(\alpha)\frac{\alpha^{H^\star}}{(\alpha + H)^H} = \frac{\alpha^{H^\star + a^\alpha - 1}}{(\alpha + H)^H}\exp\left(-\alpha b^\alpha\right). \tag{21}$$

这样便可以轻松实现 MH步骤来抽样 α。或者，我们也可以通过吉布斯步骤（类似于（20）），基于$\alpha|\tau_1, \ldots, \tau_H \sim \mathcal{G}(a^\alpha + H, b^\alpha - \frac{1}{H}\sum_{h=1}^{H} \log\ \tau_h)$，在给定 slab 概率 $\tau_1, \ldots, \tau_H$的条件下抽样 α。然而，在实践中我们发现，与从 $p(\alpha|H^\star)$进行边缘抽样相比，条件抽样的混合效果较差。

5 稀疏贝叶斯因子分析中的应用

5.1 载荷矩阵的列特异性收缩

累积收缩过程先验的一个常见应用，是在具有 $H > H_0$ 潜在因子的过拟合因子模型中，通过 $m \times H$ 因子载荷矩阵 $\boldsymbol{\beta}$ 的活跃列数量 H_0 来识别未知的因子数量，

$$\mathbf{y}_t = \boldsymbol{\beta}\,\mathbf{f}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}_m\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \quad \mathbf{f}_t \sim \mathcal{N}_H\left(\mathbf{0}, \mathbf{I}\right), \tag{22}$$

其中，$\boldsymbol{\Sigma} = \text{Diag}\left(\sigma_1^2, \ldots\ldots, \sigma_m^2\right)$ 是一个对角元素均严格为正的对角矩阵，参见 Ro˘ckov´a 和 George（2017）；Fr¨uhwirth-Schnatter 等人（2022b）以及其他许多文献。为了引入逐渐增强的列特异性收缩，并区分 $\boldsymbol{\beta}$ 中的活跃列与非活跃列，Legramanti 等人（2020）在无限贝叶斯因子分析中，将 CUSP 先验（2）应用于过拟合因子模型（22）中的 $H = \infty$。在厚尾部分，假设 $\boldsymbol{\beta}$的元素 β_{ih}服从条件高斯分布，其结构化先验方差取决于全局收缩参数 κ 和列特异性收缩参数 θ_h，

$$\beta_{ih}|\kappa, \theta_h, \sigma_i^2 \sim \mathcal{N}\left(0, \kappa\theta_h\sigma_i^2\right). \tag{23}$$

假设特异方差 σ_i^2服从逆伽马先验 $\sigma_i^2 \sim \mathcal{G}^{-1}\left(c^\sigma,\ b^\sigma\right)$，对所有$i = 1, \ldots\ldots, m$均如此。Schiavon 等人（2021）的近期研究将 CUSP 先验扩展到广义无限因子分解模型，其中因子载荷 β_{ih} 可以恰好为零，另见 Fr¨uhwirth-Schnatter 等人（2022b）。

作为示例，我们在一个含有 $H \leq H_{\max}$ 的有限过拟合模型中，考虑列特异性收缩参数 θ_h 上的有限广义CUSP先验，其中 $H_{\max} = \lfloor (m-1)/2 \rfloor$ 等于 Anderson 和 Rubin (1956) 的上界，从而确保计量经济学识别。我们对载荷矩阵的元素 β_{ih} 使用先验 (23)，并为 θ_h 引入一种比以往文献更一般的尖峰与厚尾先验。具体而言，我们假设 $h = 1, \ldots\ldots, H$ 遵循以下分层ESP先验：

$$\theta_h|S_h, \nu_0 \sim (1 - S_h)\nu_0\text{F}(2a^\theta, 2c^\theta) + S_h\text{F}(2a^\theta, 2c^\theta), \tag{24}$$

$$\text{P}(S_h = 1|\tau_h) = \tau_h, \quad \tau_h|\alpha, H \sim \mathcal{B}\left(\tfrac{\alpha}{H}, 1\right). \tag{25}$$

该列特异性方差参数 θ_h 的尖峰与厚尾先验基于 F 分布。Cadonna 等人（2020）指出，F 分布是方差参数的一种非常有用的先验。在此情境下，它被称为三重伽马先验。在尖峰部分，假设采用平移F分布 $\theta_h|S_h = 0 \sim \nu_0\text{F}(2a^\theta,\ 2c^\theta)$，其中 $\nu_0 << 1$ 充当缩减因子，使厚尾部分的累积分布函数趋近于 0。显然，该先验满足命题 2.1 的假设，其中 $\theta_0 = 0$:

$$P_{\text{slab}}(\mathbb{B}_\varepsilon(0)) = P_{\text{slab}}(\theta_h \leq \varepsilon) = P_{\text{spike}}(\theta_h \leq \frac{\varepsilon}{\nu_0}) < P_{\text{spike}}(\theta_h \leq \varepsilon) = P_{\text{spike}}(\mathbb{B}_\varepsilon(0)).$$

先验（24）相当灵活，并扩展了文献中此前提出的各种收缩先验。它可表示为逆伽马分布的层次混合：

$$\theta_h|S_h,\nu_0,b_h^\theta \sim (1-S_h)\mathcal{G}^{-1}(c^\theta,\nu_0 b_h^\theta)+S_h\mathcal{G}^{-1}(c^\theta,b_h^\theta),\quad b_h^\theta\sim\mathcal{G}(a^\theta,a^\theta/c^\theta). \tag{26}$$

随着 a^θ 增大，b_h^θ 收敛到 c^θ，且式（26）与 Legramanti 等（2020）相关。厚尾分布趋近于 $\theta_h|S_h=1\sim\mathcal{G}^{-1}(c^\theta,\ c^\theta)$，而平移尖峰分布 $\theta_h|S_h=0\sim\mathcal{G}^{-1}(c^\theta,\ c^\theta\nu_0)$ 则以期望先验为 ν_0 的连续分布取代了 Legramanti 等（2020）中的狄拉克尖峰 δ_{θ_∞}（其中 $\theta_\infty=\nu_0$）。对于 $a^\theta=1$，随着 c^θ 增大，先验（24）趋近于 Ro˘ckov´a 和 George（2017）的 Lasso 先验混合分布。最后，对于 $a^\theta=0.5$，先验（24）与 Kowal 和 Canale（2022）近期提出的先验密切相关。作为示例，我们在第 5.2 节中应用先验（24）的这三个特殊情形。

ESP 先验（定义于式（24）和（25））中的重要超参数（如 α、ν_0 和 κ）会在 MCMC 抽样过程中，借助合适的先验从数据中学习。对于 1PB 先验（25）中的超参数 α，我们遵循第 4.4 节的做法，选择 Fr¨uhwirth-Schnatter 等人（2022b）采用的伽马先验 $\alpha\sim\mathcal{G}(a^\alpha,\ b^\alpha)$。正如 Schiavon 和 Canale（2020）所指出的，混合模型（24）中的缩减因子 ν_0 可能会产生极大影响，因此必须谨慎选择。为此，我们对该参数设定均值为 $E^\nu << 1$的伽马先验，具体为 $\nu_0\sim\mathcal{G}(c^\nu,\ c^\nu/E^\nu)$。利用由式（24）和（25）推导出的混合似然，$\nu_0$ 从条件后验分布中抽样

$$p(\nu_0|\{\theta_h,\tau_h\}_h)\propto p(\nu_0)\prod_{h=1}^{H}\left[(1-\tau_h)p_{\text{spike}}(\theta_h|\nu_0)+\tau_h p_{\text{slab}}(\theta_h)\right] \tag{27}$$

并使用 MH 步骤。尖峰和厚尾密度可根据其对应的 F$(2a^\theta,\ 2c^\theta)$ 分布轻松推导：

$$\begin{aligned} p_{\text{slab}}(\theta_h) &= \frac{a^\theta}{c^\theta B(a^\theta,c^\theta)}\left(\frac{a^\theta\theta_h}{c^\theta}\right)^{a^\theta-1}\left(1+\frac{a^\theta\theta_h}{c^\theta}\right)^{-(a^\theta+c^\theta)}, \\ p_{\text{spike}}(\theta_h|\nu_0) &= \frac{1}{\nu_0}p_{\text{slab}}\left(\frac{\theta_h}{\nu_0}\right). \end{aligned} \tag{28}$$

全局收缩参数 κ 从条件逆伽马后验分布中抽样，其先验为 $\kappa\sim\mathcal{G}^{-1}(c^\kappa,b^\kappa)$

$$\kappa|\boldsymbol{\beta},\theta_1,\ldots,\theta_H,\boldsymbol{\Sigma}\sim\mathcal{G}^{-1}(c^\kappa+\frac{mH}{2},b^\kappa+\frac{1}{2}S_{\boldsymbol{\beta}}),\quad S_{\boldsymbol{\beta}}=\sum_{h=1}^{H}\frac{1}{\theta_h}\sum_{i=1}^{m}\frac{1}{\sigma_i^2}\beta_{ih}^2. \tag{29}$$

实现 ESP 先验的马尔可夫链蒙特卡罗时，一个重要步骤是对指示变量 $S_1,\ldots,S_H$进行采样，以将载荷矩阵的列划分为活跃列和非活跃列。如第 4.1 节所述，可基于 F 混合 (24) 并使用尖峰和厚尾密度 (28) 进行分类：

$$\mathrm{P}(S_h=0|q_A,\theta_h,\nu_0)\propto(1-q_A)\cdot p_{\text{spike}}(\theta_h|\nu_0),\quad \mathrm{P}(S_h=1|q_A,\theta_h)\propto q_A\cdot p_{\text{slab}}(\theta_h). \tag{30}$$

马尔可夫链蒙特卡罗程序的完整细节见算法 1。下文将讨论另一种马尔可夫链蒙特卡罗方案: 在给定 $\boldsymbol{\beta}_h$ 和 b_h^θ 的条件下对 S_h 进行采样，同时对 θ_h 进行边缘化。

算法 1(F- 分类)。 一次马尔可夫链蒙特卡罗估计迭代包含以下抽样步骤:

(1) 从 $p(\boldsymbol{\beta}, \sigma_1^2, \ldots, \sigma_m^2 | \theta_1, \ldots, \theta_H, \kappa, f_1, \ldots, f_n, y)$ 中抽取模型参数 $(\boldsymbol{\beta}, \sigma_1^2, \ldots, \sigma_m^2)$。对于 $t = 1, \ldots, n$，从 $p(f_t | \boldsymbol{\beta}, \sigma_1^2, \ldots, \sigma_m^2, y)$ 中抽取潜在因子 f_t。*(2)* 对 $\log \nu_0$ 使用标准随机游走 *MH* 步骤，从式 *(27)* 给出的后验分布 $p(\nu_0 | \{\theta_h, \tau_h\}_{h=1}^H)$ 中抽取 ν_0。对于 $h = 1, \ldots, H$，使用 $q_A = \alpha/(\alpha + H)$ 从离散后验分布 *(30)* 中抽取 S_h。使用标准随机游走 *MH* 步骤，对 $\log \alpha$ 从 $p(\alpha | H^\star)$ 中抽取 α，其中 $p(\alpha | H^\star)$ 见式 *(21)*。对于 $h = 1, \ldots, H$，从贝塔分布 *(17)* 中抽取 $\tau_h | S_h, \alpha$，其中 $a_0 = \alpha/H$ 和 $b_0 = 1$。根据式 *(31)* 抽取 $b_h^\theta | \theta_h, \nu_0, S_h$ 和 $\theta_h | \boldsymbol{\beta}_h, \boldsymbol{\Sigma}, b_h^\theta, \nu_0, S_h$。从逆伽马分布 *(29)* 中抽取 $\kappa | \theta_1, \ldots, \theta_H, \boldsymbol{\beta}$。

算法 1 的第 (1) 部分包含贝叶斯因子分析中的标准步骤，参见例如 Fr¨uhwirth- Schnatter 等人（2022b）。第 (2) 部分包含实现式 (24) 和 (25) 中引入的 ESP 先验所需的所有步骤。条件后验分布 $\theta_h | \boldsymbol{\beta}_h$， $\boldsymbol{\Sigma}$， b_h^θ， ν_0， S_h 和 $b_h^\theta | \theta_h$， ν_0， S_h 由式 (26) 推导得出:

$$\theta_h | \boldsymbol{\beta}_h, \boldsymbol{\Sigma}, b_h^\theta, \nu_0, S_h \sim \mathcal{G}^{-1}(c^\theta + \frac{m}{2}, \nu_0^{1-S_h} b_h^\theta + \frac{1}{2\kappa} \sum_{i=1}^m \frac{\beta_{ih}^2}{\sigma_i^2}), \tag{31}$$

$$b_h^\theta | \theta_h, \nu_0, S_h \sim \mathcal{G}(a^\theta + c^\theta, \frac{a^\theta}{c^\theta} + \frac{\nu_0^{1-S_h}}{\theta_h}).$$

为改善混合效果，每个周期结束时都会执行一个增强步骤，其中涉及 θ_1、……、θ_H和 κ，如 Fr¨uhwirth- Schnatter 等人（2022b）所述。还可以采用另一种马尔可夫链蒙特卡罗方案，在抽样指示变量时对 θ_h进行边缘化。 S_1、……、S_H。表示式 (26) 可用于推导在给定 b_h^θ 时， $\boldsymbol{\beta}$ 的第 h 列 $\boldsymbol{\beta}_h$ 的条件尖峰与厚尾分布，如下列 t-分布所示:

$$p_{\text{spike}}(\boldsymbol{\beta}_h | b_h^\theta, \nu_0, \kappa, \boldsymbol{\Sigma}) \sim t_{2c^\theta}(\mathbf{0}, \frac{\nu_0 \kappa b_h^\theta}{c^\theta} \boldsymbol{\Sigma}), \qquad p_{\text{slab}}(\boldsymbol{\beta}_h | b_h^\theta, \kappa, \boldsymbol{\Sigma}) \sim t_{2c^\theta}(\mathbf{0}, \frac{\kappa b_h^\theta}{c^\theta} \boldsymbol{\Sigma}). \tag{32}$$

这些密度可用于关于 θ_h 边际化的分类:

$$\begin{aligned} \mathrm{P}(S_h = 0 | q_A, \boldsymbol{\beta}_h, \nu_0, b_h^\theta, \kappa, \boldsymbol{\Sigma}) &\propto (1 - q_A) \cdot p_{\text{spike}}(\boldsymbol{\beta}_h | b_h^\theta, \nu_0, \kappa, \boldsymbol{\Sigma}), \\ \mathrm{P}(S_h = 1 | q_A, \boldsymbol{\beta}_h, b_h^\theta, \kappa, \boldsymbol{\Sigma}) &\propto q_A \cdot p_{\text{slab}}(\boldsymbol{\beta}_h | b_h^\theta, \kappa, \boldsymbol{\Sigma}), \end{aligned} \tag{33}$$

并从条件后验分布中采样 ν_0

$$p(\nu_0 | \{\boldsymbol{\beta}_h, b_h^\theta, \tau_h\}_h, \kappa, \boldsymbol{\Sigma}) \propto p(\nu_0) \prod_{h=1}^H \left[(1 - \tau_h) p_{\text{spike}}(\boldsymbol{\beta}_h | b_h^\theta, \nu_0, \cdot) + \tau_h p_{\text{slab}}(\boldsymbol{\beta}_h | b_h^\theta, \cdot) \right] \tag{34}$$

使用 MH 步骤。此外，由于对 θ_h 而非 b_h^θ 进行边缘化，$\theta_h|b_h^\theta$、$\cdot$和 $b_h^\theta|\theta_h$、$\cdot$的采样顺序与算法 1 相比发生了反转。该 MCMC 程序的完整细节见算法 2。

算法 2(t-c　　分类)。 一次马尔可夫链蒙特卡罗估计迭代包含以下抽样步骤：

(1) 与算法 *1* 相同。*(2)* 使用针对$\log \nu_0$的标准随机游走 *MH* 步骤，从式 *(34)* 给出的后验分布中抽取 ν_0。对于 $h = 1, \ldots, H$，使用 $q_A = \alpha/(\alpha + H)$ 从离散后验分布 *(33)* 中抽取 S_h。按照算法 *1* 抽取 $\alpha|H^\star$ 和 $\tau_h|S_h, \alpha, h = 1, \ldots, H$。从式 *(31)* 中抽取 $\theta_h|\boldsymbol{\beta}_h, \boldsymbol{\Sigma}, b_h^\theta, \nu_0, S_h$ 和 $b_h^\theta|\theta_h, \nu_0, S_h$。按照算法 *1* 抽取 $\kappa|\theta_1, \ldots, \theta_H, \boldsymbol{\beta}$。

一般而言，我们发现算法 1 的混合效果优于算法 2；算法 2 往往会停留在 H_0 的真实值处，从而夸大后验集中度；示例见第 5.2 节。

如第 3 节所述，ESP 先验 (24) 和 (25) 实现了递增收缩，却不会隐含地强制各列按特定顺序排列。因此，算法 1 和 2 不会对载荷矩阵的列施加任何排序限制。在马尔可夫链蒙特卡罗过程中，列的顺序保持任意，这大大简化了抽样。不过，在马尔可夫链蒙特卡罗过程中，仍可根据 $S_1, \ldots, S_H$ 的后验抽样，求得活跃列的数量 $H^\star$，因为函数泛函 (19) 对列的顺序保持不变（后验抽样的其他几个函数泛函也具有这一性质）。递增收缩在马尔可夫链蒙特卡罗抽样的后处理中变得明显。首先，对于每次抽样得到的、无序的 slab 概率 $\tau_1, \ldots, \tau_H$，我们求出其降序次序统计量 $\tau_{(1)} > \cdots > \tau_{(H)}$（在算法 1 或算法 2 的第 (2) 部分中抽样），并利用相应的排列 $\rho(1) \ldots, \rho(H)$ 对载荷矩阵的列重新排序。由此可得到 ESP 先验的广义 CUSP 表示，其中序列 $\pi_h = 1 - \tau_{(h)}$ 包含递增的尖峰概率，相应的列特定参数为 $(\theta_1^\star, \ldots, \theta_H^\star) = (\theta_{\rho(1)}, \ldots, \theta_{\rho(H)})$。随着列索引 h 增大，$\theta_h^\star$ 和 π_h 的边际后验分布分别越来越趋近于零和一，图 1 展示了这一现象。

5.2 一个说明性模拟研究

我们进行了与 Legramanti 等（2020）类似的模拟研究，并考虑了三种不同的 (m, H_0) 组合，分别为 (20，5)、(50，10) 和 (100，15)。对于每种 (m, H_0) 组合，我们从高斯因子模型 (1) 中抽取 25 个包含 $n = 100$ 个观测的数据集。除了 Legramanti 等（2020）采用的稠密设定（其中载荷矩阵 $\boldsymbol{\beta}_0$ 的所有元素 β_{ih} 均不受约束）

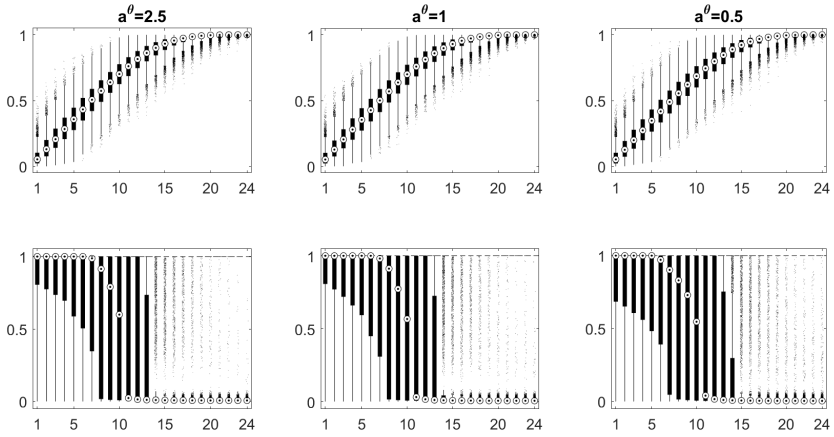

图 1：在稠密情景下模拟的数据集，其中 $(m,\ H_0) = (50,\ 10)$。三种 ESP 先验的 CUSP 表示（从左到右），展示了递增的尖峰概率 π_h（上）及对应的列特定收缩参数 $\theta_h^\star$（下）的后验抽样箱线图，列索引 $h = 1, \ldots\ldots, 24$ 递增。

且独立地从 $\mathcal{N}(0,\ 1)$ 中抽取之外，我们还考虑稀疏 设定，其中所有 $\beta_{ih} = 0$ 中有 30%，其余 70% 则从标准正态分布中抽取。在全部六种情形下，$\boldsymbol{\Sigma}_0 = \mathbf{I}$。

过拟合因子模型中的活跃列最大数量 $H = \max(H_{\max},\ 30)$ 随 m 增加；出于计算方面的原因，$m = 100$ 时该数量限制为 $H = 30$。对于 θ_h 上的 ESP 先验，我们研究了 (24) 中的三种不同尖峰与厚尾先验：带有 $a^\theta = 2.5$的 F 混合、正则化 Lasso 混合（$a^\theta = 1$）和正则化 horseshoe 混合（$a^\theta = 0.5$）；而 $c^\theta = 2.5$ 则与 Legramanti 等（2020）和 Kowal 和 Canale（2022）相同。对于 (25) 中 τ_h 上的先验，我们在伽马先验 $\alpha \sim \mathcal{G}(6,\ 2)$ 下，根据数据学习 α。这一选择意味着较大的先验概率 P($\alpha << H$)，并对活跃列数量 $H^\star$ 引入先验收缩。其他超参数选择为 $c^\sigma = 2.5,\ b^\sigma = 1.5$、$c^\kappa = b^\kappa = 5$、$c^\nu = 10$ 和 $E^\nu = 0.01$，相应的先验期望分别为 $\mathbb{E}(\sigma_i^2) = 1$、$\mathbb{E}(\kappa^{-1}) = 1$ 和 $\mathbb{E}(\nu_0) = 0.01$。算法 1 运行 10,000 次迭代，预热期为 5,000 次，初始模型包含 $H^\star = 3$ 个活跃列。计算在一台配备英特尔酷睿 i5-8265U 处理器（1.60–1.80 GHz）的笔记本电脑上使用 MATLAB 2020 完成。

对于25个模拟数据集中的每一个，我们通过蒙特卡罗估计计算以下统计量，以评估 $H^\star$ 作为真实因子数 H_0 估计量ˆ的可靠性：我们考察后验分布 $p(H^\star|\mathbf{y})$的众数 $H^\star$ 以及后验密度值 $p(H^\star = H_0|\mathbf{y})$的大小。为评估对真实值的估计准确性

表1：在 $a^\theta = 2.5$ (F)、 $a^\theta = 1$(L)、 $a^\theta = 0.5$ (H)、 $c^\theta = 2.5$和 $\alpha \sim \mathcal{G}$(6,2)下，ESP先验（24）在6种不同数据场景中的表现。M 和Q 列给出了各种统计量在模拟数据集上的中位数、5%分位数和95%分位数。CUSP先验（C）的结果基于Legramanti等（2020）的表1，其中Q 列显示四分位距。最后一列显示运行时间（单位为CPU秒）的中位数 $\tilde{s}$ 。

			$\hat{H}^\star$		$p(H^\star = H_0 \mid \mathbf{y})$		MSE_Ω		
$(m,\ H_0)$		先验	M	Q	M	Q	M	Q	$\tilde{s}$
(20,5)	稠密	F	5	(5,5)	0.96	(0.87,0.98)	0.78	(0.54,1.09)	50.2
		L	5	(5,5)	0.85	(0.64,0.90)	0.79	(0.56,1.46)	50.4
		H	5	(5,5)	0.66	(0.47,0.71)	0.87	(0.50,1.21)	51.1
		C	5	0	-	-	0.75	0.29	310.8
	稀疏	F	5	(5,5)	0.92	(0.56,0.97)	0.50	(0.30,0.75)	51.1
		L	5	(5,5)	0.80	(0.62,0.87)	0.46	(0.23,0.60)	51.2
		H	5	(5,5)	0.55	(0.40,0.68)	0.51	(0.27,1.16)	51.9
(50,10)	稠密	F	10	(10,10)	0.98	(0.94,0.99)	2.23	(1.60,2.95)	268.3
		L	10	(10,10)	0.86	(0.81,0.89)	2.17	(1.70,3.16)	271.8
		H	10	(10,10)	0.57	(0.51,0.61)	2.32	(1.75,3.26)	273.6
		C	10	0	-	-	2.25	0.33	716.2
	稀疏	F	10	(10,10)	0.97	(0.95,0.98)	1.16	(0.80, 1.61)	261.5
		L	10	(10,10)	0.81	(0.72,0.87)	1.20	(0.87,1.66)	266.0
		H	10	(10,10)	0.52	(0.44,0.57)	1.17	(0.85,1.97)	266.8
(100,15)	稠密	F	15	(15,15)	0.99	(0.98,0.99)	3.59	(3.08,4.50)	1219.0
		L	15	(15,15)	0.88	(0.86,0.90)	3.91	(3.35,4.35)	1219.8
		H	15	(15,15)	0.57	(0.53,0.61)	3.81	(3.26,5.02)	1252.7
		C	15	0	-	-	3.76	0.4	2284.9
	稀疏	F	15	(15,15)	0.97	(0.96,0.99)	1.93	(1.57,2.29)	1188.1
		L	15	(15,15)	0.84	(0.80,0.89)	1.97	(1.63, 2.34)	1193.8
		H	15	(15,15)	0.52	(0.40,0.56)	2.10	(1.63,2.35)	1216.7

数据的协方差矩阵 $\mathbf{\Omega}_0 = \boldsymbol{\beta}_0\boldsymbol{\beta}_0^\top + \mathbf{\Sigma}_0$ 通过过拟合模型 (22) 所隐含的协方差矩阵 $\mathbf{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}^\top + \mathbf{\Sigma}$ ，我们考虑如下定义的均方误差 (MSE)：

$$\text{MSE}_\Omega = \sum_i \sum_{\ell \le i} \mathbb{E}((\Omega_{i\ell} - \Omega_{0,i\ell})^2 \,|\mathbf{y})/(m(m+1)/2).$$

表1报告了，f 对于全部 18 种可能的数据场景和 ESP 先验，这些统计量在全部 25 个模拟数据集上的中位数、5% 分位数以及 e

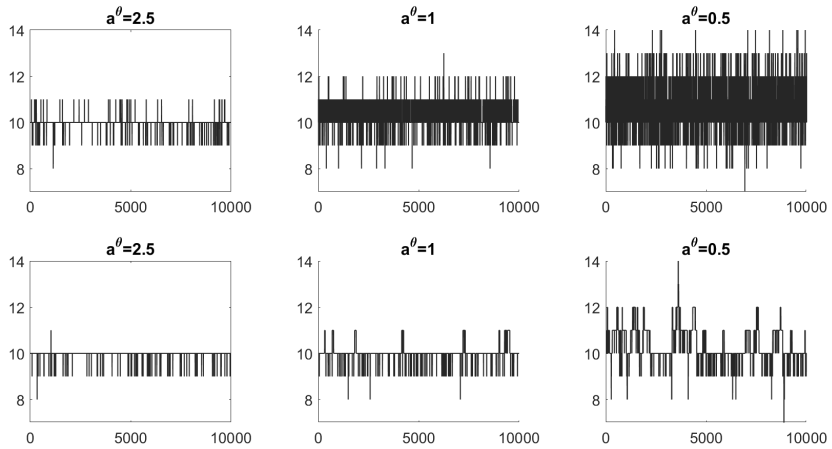

图2：在稠密情景下模拟的数据集，其中 $(m,\ H_0) = (50,\ 10)$。对于带有 $a^\theta = 2.5$ （左）、$a^\theta = 1$ （中）、 $a^\theta = 0.5$ （右）、 $c^\theta = 2.5$ 和 $\alpha \sim \mathcal{G}(6,\ 2)$ 的 ESP 先验，使用算法 1（上）和算法 2（下）得到的 $H^\star$ 10,000 个后验抽样。

95% 分位数。关于协方差矩阵 $\mathbf{\Omega}_0$,，三种 ESP 先验的均方误差大致相同，且随着 H_0 增大而增大；对于稀疏载荷矩阵，其均方误差显著小于稠密载荷矩阵。三种 ESP 先验在根据后验众数 $H^\star$ ^ 重新估计 H_0 方面同样成功。有趣的是，在三种 a^θ 取值下，真实值 H_0 处的后验集中度存在显著差异。在全部六种数据场景中，F 混合且 $a^\theta = 2.5$ 时的后验密度值 $p(H^\star = H_0|$ y) 接近 1，即使在稀疏设定下也是如此。对于另外两种 a^θ 选择， $p(H^\star = H_0|$y) 的取值显著小于 1，尤其是正则化 horseshoe 混合。

如第 5.1 节末尾所述，无序 slab 概率 τ_1、.....、 τ_H 的降序次序统计量 $\tau_{(1)} > \cdots > \tau_{(H)}$ 可用于获得各种 ESP 先验的 CUSP 表示。为获得更多见解，本文报告了在稠密情景下模拟的数据集的详细结果，其中 $(m,\ H_0) = (50,\ 10)$。图1展示了所有三种 ESP 先验下，CUSP 表示中递增的尖峰概率 $\pi_h = 1 - \tau_{(h)}$ 以及相应列特异性收缩参数 $\theta_h^\star$ 的后验抽样箱线图。由于 ESP 先验的隐式 CUSP 性质，尖峰概率 π_h 的后验分布逐渐被拉向 1，而 $\theta_h^\star$ 的后验分布则随着 h 增大而被拉向 0。在所有三种 ESP 先验下，数据中的信息在真实值 $h = H_0$ =10处诱导出活跃列与非活跃列之间明显的后验间隔。

关于马尔可夫链蒙特卡罗性能，算法 1 对所有已识别参数都表现出良好的混合性质。

在未对 10,000 个后验抽样进行任何稀疏化的情况下，所有 25 个模拟数据集中的 $\log|\boldsymbol{\Omega}|$ 和 $\|\boldsymbol{\Omega}^{-1}\|_F$ 的中位数有效抽样率平均分别为 27.7% 和 14.4%，对应的平均中位数有效样本量（ESS）分别为 2768 和 1435。为进一步说明，图2展示了针对一个在稠密情景下模拟的数据集 $(m, H_0) = (50,10)$，使用算法 1 和算法 2、基于全部三种 ESP 先验得到的 $H^\star$ 的 10,000 个后验抽样。显然，算法 1 使用 θ_h 的 F 混合来区分活跃列与非活跃列，其混合表现远优于算法 2；后者利用载荷矩阵列 $\boldsymbol{\beta}_h$ 的边缘化 t-混合来实现这一目的。

图2中 $H^\star$ 后验抽样的变化反映了后验分布 $p(H^\star|\mathbf{y})$的集中程度。如前所述，对于正则化 horseshoe 混合，后验密度值 $p(H^\star = H_0 = 10|\mathbf{y})$远小于 1（再次参见表1）。在算法 1 下，相应的后验抽样在离散后验分布 $p(H^\star|\mathbf{y})$上表现出极佳的混合，这正是 Kowal 和 Canale（2022）最初建议采用该先验的主要动机。然而，这是以较低的后验集中度为代价的。对于正则化 Lasso 混合，后验集中度则更为明显（再次参见表1）。尽管如此，在算法 1 下，后验抽样仍能在后验分布 $p(H^\star|\mathbf{y})$上快速移动。对于 $a^\theta = 2.5$，算法 1 得到的后验分布 $p(H^\star|\mathbf{y})$高度集中；这不仅适用于此特定数据集，也适用于大多数其他数据集（再次参见表1）。

Kowal 和 Canale（2022）提出了一个合理的问题：如此强的后验集中度是否是采样器混合不佳所致？对于图2中的具体示例，算法 2证实了在 $a^\theta = 2.5$ 下几乎完美的后验集中度。然而，在450个模拟数据集中，我们发现了许多情况，尤其是对于 $a^\theta = 1$ 和 $a^\theta = 0.5$，算法 2迅速从包含三个活跃列的初始模型转移到 H_0的真实值，随后却卡在 H_0，与算法 1相比产生了过于乐观的后验集中度（算法 1在这些情况下也混合良好）。

最后，图3展示了稠密情景下（m，H_0）=（50，10）中，尖峰与厚尾先验在 θ_h 上的 1BP先验强度参数 α在 a^θ 三种取值下的后验抽样。关于 α 的后验推断对 a^θ 的选择相当稳健，并强烈支持 α取值显著小于 $H = 24$。

与其他先验的比较。选择 $\alpha = H$ 对应于 Zhao 等（2016）采用的均匀先验 $\tau_h \sim \mathcal{U}[0, 1]$，而在固定 $\alpha = H$ 的情况下，可使用算法 1 的简化版本，在该先验下进行后验推断。表 2 显示了在尖峰与厚尾先验下，对 θ_h 采用尖峰与厚尾先验时，稀疏数据场景与超参数 a^θ 的所有 9 种组合下，各种统计量如何随该先验而变化。值得注意的是，数据是否具有足够的信息量，能够克服均匀先验对尖峰概率先验分布的强烈影响（尖峰概率会远离 1，参见

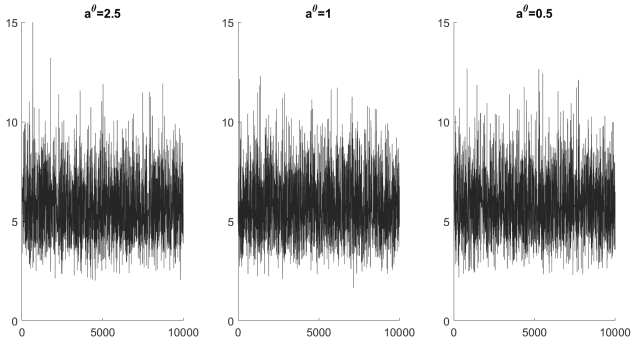

图3：在稠密情景下模拟的数据集，参数为 $(m, H_0) = (50,10)$。针对 α，ESP 先验的 10,000 次后验抽样，其中 $a^\theta = 2.5$（左侧）、$a^\theta = 1$（中间）、$a^\theta = 0.5$（右侧）、$c^\theta = 2.5$和 $\alpha \sim \mathcal{G}$（6,2）。

表 2：在三个稀疏数据场景下，尖峰与厚尾先验（24）分别采用 $a^\theta = 2.5$ (F)、$a^\theta = 1$ (L)、$a^\theta = 0.5$ (H)，以及在均匀先验 $\tau_h \sim \mathcal{U}[0,1]$ 下采用 $c^\theta = 2.5$ 时的表现。列 M 和 Q 分别展示了模拟数据集上各种统计量的中位数、5% 分位数和 95% 分位数。

		$\hat{H}^\star$		$p(H^\star = H_0 \mid \mathbf{y})$		MSE_Ω	
(m, H_0)	先验	M	Q	M	Q	M	Q
(20,5)	F	5	(5,5)	0.91	(0.57,0.97)	0.48	(0.31,0.96)
	L	5	(5,5)	0.68	(0.28,0.78)	0.42	(0.28,0.68)
	H	5	(5,6)	0.41	(0.32,0.51)	0.52	(0.21,0.89)
(50,10)	F	10	(10,10)	0.96	(0.89,0.98)	1.14	(0.90,1.58)
	L	10	(10,10)	0.53	(0.45,0.61)	1.21	(0.89,1.65)
	H	12	(11,12)	0.13	(0.08,0.16)	1.27	(0.99,1.87)
(100,15)	F	15	(15,15)	0.98	(0.98,0.99)	1.97	(1.69,2.61)
	L	15	(15,15)	0.62	(0.56,0.67)	2.10	(1.64,2.47)
	H	17	(16,17)	0.13	(0.10,0.16)	2.00	(1.68,2.41)

第 4.4 节）取决于所选尖峰与厚尾先验的设定。在采用参数为 $a^\theta = 2.5$ 的 F 混合先验以及正则化Lasso混合先验时，我们仍能找回真实因子数，但后验集中度不如之前。另一方面，采用正则化马蹄混合先验时，真实因子数会被系统性地高估，尤其是在两个较大的模型中。

此外，为了将有限 ESP 先验与原始 CUSP 先验进行比较，我们在表1中复现了 Legramanti 等（2020）针对稠密因子模型报告的表1中的部分性能指标。所有有限 ESP 先验在关于 $H^\star$ 的推断方面均与 CUSP 先验具有相同的统计性能，在均方误差$_\Omega$方面也非常相似。然而，有限 ESP 先验显著缩短了运行时间，这是因为抽样二元指示变量 $S_1, \ldots, S_H$ 取代了抽样类别指示变量 $z_1, \ldots, z_H$。此外，对于上述的两个 **Ω** 统计量，我们获得了比 Legramanti 等（2020）更高的有效抽样量；在他们的研究中，2,000 个稀疏抽取样本的有效样本量中位数，对于一个略有不同的 **Ω**统计量，平均为 368。这种抽样效率的提升源于算法 1 中的部分边缘化：我们在不以 slab 概率 $\tau_1, \ldots, \tau_H$为条件的情况下，抽样强度参数 α 和指示变量 $S_1, \ldots, S_H$，参见第 4.1 节。

6 结论

本文讨论了能够自动对一系列参数施加递增收缩的收缩先验。我们的主要动机来自贝叶斯因子分析，在该分析中，随着列索引增加，会对载荷矩阵施加递增收缩，从而能够对未知因子维度进行统计推断。

我们简要回顾了 Legramanti 等（2020）的 CUSP 先验。这是一种尖峰与厚尾先验，其中尖峰概率随机递增，并由狄利克雷过程先验的打破棍表示构造而成。作为第一项贡献，我们将该先验扩展为涉及任意打破棍表示的广义 CUSP 先验。该先验涵盖了文献中较早提出的若干先验，这些先验涉及基于贝塔分布的各种打破棍表示 Ro˘ckov´a and George（2017）；Heaukulani 和 Roy（2020）；Ohn 和 Kim（2022）；Fr¨uhwirth-Schnatter 等（2022b）；Kowal 和 Canale（2022）。作为第二项贡献，我们证明，在贝叶斯推断的许多应用领域中广受欢迎且被广泛使用的可交换尖峰与厚尾收缩（ESP）先验，可以表示为有限广义 CUSP 先验。CUSP 表示可以由厚尾概率的降序次序统计量轻松推导得到。

对一系列参数使用 ESP 先验非常方便：该先验对参数的排列顺序不变，同时能够隐式地施加递增收缩，而无需对厚尾概率施加显式的顺序约束。尤其是，这使得我们能够在 ESP 先验下设计高效的马尔可夫链蒙特卡罗抽样器，并在后处理中推导出 CUSP 表示。相比之下，直接在 CUSP 表示的顺序约束下进行抽样更具挑战性；对于无限维模型，通常必须采用带有 $H < \infty$ 的截断 CUSP 先验。若改用具有较大 H 的 ESP 先验，并对参数 θ_h 采用与无限 CUSP 先验相同的尖峰与厚尾分布，则会产生类似的递增收缩，同时大大简化分类过程，将其化简为对H 个二元指示变量进行抽样，而非对具有 H 个类别的 H 个分类变量进行抽样。

将其应用于稀疏贝叶斯因子分析，展现了本文研究结果的实用价值。本文提出了一种基于三重伽马先验的全新可交换尖峰与厚尾收缩先验，见 Cadonna 等人（2020）。在贝叶斯因子分析中，该 ESP 先验会使载荷矩阵的列呈现递增收缩。模拟研究表明，该先验有助于估计未知的因子数量。此处稀疏贝叶斯因子分析应用的重点在于列稀疏性，但正如引言所述，逐元素稀疏性也是因子分析中的另一项常见目标。结合这两种方法是一个值得进一步研究的方向，Fr¨uhwirth-Schnatter 等人（2022b）对此进行了探讨。

附录

我们证明，有限1PB先验（14）可表示为式（5）和（6）中的有限CUSP先验，其中折棒 ν_h、$h = 1$、.....、H 是相互独立的贝塔随机变量序列，

$$\nu_h \sim \mathcal{B}\left(1, \alpha\frac{H-h+1}{H}\right). \tag{35}$$

等价地，降序的厚尾概率可表示为：

$$\tau_{(h)} = 1 - \pi_h = \prod_{\ell=1}^{h}(1-\nu_\ell) = \prod_{\ell=1}^{h}\nu_\ell^\star = \tau_{(h-1)}\nu_h^\star, \tag{36}$$

其中$\nu_h^\star$ $i.i.d.$ $\mathcal{B}\left(\alpha\frac{H-h+1}{H}, 1\right)$。容易证明，对于所有 h， $\mathbb{E}(\pi_h)$满足以下递推关系。由（36）可得

$$\mathbb{E}(\tau_{(h)}) = \mathbb{E}(\tau_{(h-1)})\mathbb{E}(\nu_h^\star) = \mathbb{E}(\tau_{(h-1)})C_h, \qquad C_h = \frac{\alpha(1-\frac{h-1}{H})}{\alpha(1-\frac{h-1}{H})+1} < 1.$$

因此， $\mathbb{E}(\tau_{(h)})$是一个递减序列，进而 $\mathbb{E}(\pi_h) = 1 - \mathbb{E}(\tau_{(h)})$递增。由此立即可知，当 H 趋于无穷大时，先验（14）收敛到Legramanti 等（2020）提出的强度参数为 α 的CUSP先验，因为 $\lim_{H\to\infty}\frac{H-h+1}{H} = 1$。为证明（35）和（36），我们遵循Teh 等（2007）的方法，仔细考察降序次序统计量$\tau_{(1)} > \cdots > \tau_{(H)}$的分布。式（14）中的无序厚尾概率 τ_h 分别具有以下概率密度函数和累积分布函数：

$$p(\tau_h) = \frac{\alpha}{H}\tau_h^{\frac{\alpha}{H}-1}, \quad F_{\tau_h}(\tau) = \mathrm{P}(\tau_h \le \tau) = \tau^{\frac{\alpha}{H}}.$$

令 T 为任意自然数。首先证明，对于任意一列 T 个独立同分布的随机变量 $X_h \sim \mathcal{B}\left(\frac{\alpha}{H}, 1\right)$，其最大值 $X_{\max} = \max(X_1, \ldots, X_T)$ 服从 $\mathcal{B}\left(\frac{T\alpha}{H}, 1\right)$-分布：

$$F_{X_{\max}}(x) = \prod_{h=1}^{T}\mathrm{P}(X_h \le x) = (x^{\frac{\alpha}{H}})^T = x^{\frac{T\alpha}{H}}. \tag{37}$$

由 $T = H$，根据 (37) 可知，在先验 (14) 下，最大顺序统计量 $\tau_{(1)} = \max(\tau_1, \ldots, \tau_H)$ 服从 $\tau_{(1)} \sim \mathcal{B}(\alpha,\ 1)$。给定顺序统计量 $\tau_{(1)}, \ldots, \tau_{(h-1)}$ 后，其余未排序 slab 概率 $\tau_h (\tau_h < \tau_{(h-1)})$ 的取值范围显然受限于 $[0, \tau_{(h-1)}]$，相应的累积分布函数为：

$$F_{\tau_h}(\tau) = \mathrm{P}(\tau_h \leq \tau | \tau_h < \tau_{(h-1)}) = \frac{\int_0^{\tau} \frac{\alpha}{H} \tau_h^{\frac{\alpha}{H}-1} \, d\,\tau_h}{\int_0^{\tau_{(h-1)}} \frac{\alpha}{H} \tau_h^{\frac{\alpha}{H}-1} \, d\,\tau_h} = \frac{\tau^{\frac{\alpha}{H}}}{(\tau_{(h-1)})^{\frac{\alpha}{H}}} = \left(\frac{\tau}{\tau_{(h-1)}}\right)^{\frac{\alpha}{H}}.$$

因此，所有小于 $\tau_{(h-1)}$ 的 $H - h + 1$ 厚尾概率 τ_h 相互独立，并可表示为$\tau_h = X_h \tau_{(h-1)}$，其中 $X_h \sim \mathcal{B}\left(\frac{\alpha}{H}, 1\right)$。这可直接由 $-$

$$F_{X_h}(x) = \mathrm{P}(X_h \leq x) = \mathrm{P}(\tau_h \leq x\tau_{(h-1)}) = \left(\frac{x\tau_{(h-1)}}{\tau_{(h-1)}}\right)^{\frac{\alpha}{H}} = x^{\frac{\alpha}{H}}.$$

给定顺序统计量 $\tau_{(h-1)}$，顺序统计量 $\tau_{(h)}$ 可推导为 $\tau_{(h)} = \tau_{(h-1)}\nu_h^{\star}$，其中 $\nu_h^{\star} = \max_{\ell:\tau_\ell < \tau_{(h-1)}} X_h$。利用 (37) 和 $T = H - h + 1$，可得：

$$\tau_{(h)} = \tau_{(h-1)}\nu_h^{\star}, \qquad \nu_h^{\star} \sim \mathcal{B}\left(\alpha\frac{H-h+1}{H}, 1\right).$$

由此可直接得出，排序后的尖峰概率 $\pi_h = 1 - \tau_{(h)}$ 和排序后的厚尾概率 $\pi_h^{\star} = \tau_{(h)}$ 分别可表示为 (35) 和 (36) 中的形式。

参考文献

Anderson, T. W. 和 H. Rubin（1956）。因子分析中的统计推断。载于 第三届伯克利数学统计与概率研讨会论文集，第五卷，第 111–150 页。

Avalos-Pacheco, A., D. Rossell, and R. S. Savage (2022). 使用贝叶斯潜在因子回归整合异质性大型数据集。《贝叶斯分析》17, 33–66.

Bhattacharya, A. and D. Dunson (2011). 稀疏贝叶斯无限因子模型。《生物统计学》98, 291– 306.

Cadonna, A., S. Fr¨uhwirth-Schnatter, and P. Knaus (2020). 三倍伽马——一种统一的收缩先验，用于稀疏状态空间模型和时变参数模型中的方差估计与变量选择。《Econometrics 8》, 20.

Carvalho, C. M., J. Chang, J. E. Lucas, J. Nevins, Q. Wang, and M. West (2008). 高维稀疏因子建模：在基因表达基因组学中的应用。《美国统计学会杂志》103, 1438–1456.

Conti, G., S. Fr¨uhwirth-Schnatter, J. J. Heckman, 和 R. Piatek（2014）。贝叶斯探索性因子分析。《计量经济学杂志》*183*，31–57。

De Vito, R., R. Bellio, L. Trippa, 和 G. Parmigiani（2021）。面向高通量生物数据的贝叶斯多研究因子分析。《应用统计年鉴》*15*，1723 – 1741。Durante, D.（2017）。关于乘法伽马过程的简述。《统计学与概率通讯》*122*，198–204。Fr¨uhwirth-Schnatter, S., D. Hosszejni, 和 H. Lopes（2022a）。计数何时成立——基于GLT结构的因子模型计量经济学识别。*arXiv 2301.06354*。Fr¨uhwirth-Schnatter, S., D. Hosszejni, 和 H. F. Lopes（2022b）。因子数量未知时的稀疏有限贝叶斯因子分析。*arXiv 2301.06459*。Fr¨uhwirth-Schnatter, S. 和 H. Lopes（2010）。因子数量未知时的简约贝叶斯因子分析。研究报告，芝加哥大学布斯商学院。Ghahramani, Z., T. L. Griffiths, 和 P. Sollich（2007）。贝叶斯非参数潜在特征模型（含讨论和回应）。载于 J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, 和 M. West（编），《贝叶斯统计学*8*》，第1–25页。牛津：牛津大学出版社。Heaukulani, C. 和 D. M. Roy（2020）。吉布斯型印度自助餐过程。《贝叶斯分析》*15*，683–710。Hosszejni, D. 和 S. Fr¨uhwirth-Schnatter（2022）。揭开谜底！二部图揭示了稀疏因子分析中的可识别性。*arXiv2211.00671*。Kastner, G.（2019）。高维稀疏贝叶斯时变协方差估计。《计量经济学杂志》*210*，98–115。Kaufmann, S. 和 C. Schuhmacher（2019）。采用与阶数无关的事后识别的稀疏动态因子模型贝叶斯估计。《计量经济学杂志》*210*，116–134。Kowal, D. R. 和 A. Canale（2022）。采用贝叶斯秩选择的半参数函数因子模型。*arXiv2108.02151*。Legramanti, S., D. Durante, 和 D. B. Dunson（2020）。无限因子分解的贝叶斯累积收缩。《生物统计学》*107*，745–752。Murphy, K., C. Viroli, 和 I. C. Gormley（2020）。无限因子分析器的无限混合。《贝叶斯分析》*15*，937–963。Ohn, I. 和 Y. Kim（2022）。高维稀疏因子模型中因子维数的后验一致性。《贝叶斯分析》*17*，491–514。

Ro˘ckov´a, V. 和 E. I. George（2017）。通过自动旋转实现稀疏性的快速贝叶斯因子分析。美国统计学会会刊*111*，1608–1622。Schiavon, L. 和 A. Canale（2020）。关于无限因子模型的截断准则。*Stat 9*，e298。Schiavon, L., A. Canale 和 D. B. Dunson（2021）。广义无限因子分解模型。*arXiv 2103.10333*。Sethuraman, J.（1994）。狄利克雷先验的构造性定义。《中华统计学志》*4*，639–650。Teh, Y. W., D. G¨or¨ur 和 Z. Ghahramani（2007）。印度自助餐过程的折棍构造。载于 M. Meila 和 X. Shen（编），第十一届人工智能与统计国际会议论文集，机器学习研究论文集第2卷，波多黎各圣胡安，第556–563页。West, M.（2003）。「大 p、小 n」范式下的贝叶斯因子回归模型。载于 J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith 和 M. West（编），《贝叶斯统计7》，第733–742页。牛津：牛津大学出版社。Zhao, S., C. Gao, S. Mukherjee 和 B. E. Engelhardt（2016）。具有结构化稀疏性的贝叶斯群组因子分析。《机器学习研究杂志》*17*，1–47。