arXiv:2601.07864v1 [stat.ME] 10 Jan 2026

采用尖峰与宽尾先验进行完全贝叶斯变量选择的对称随机扫描塌缩吉布斯采样器

Mengta Chung

台湾淡江大学管理科学系

mtc@gms.tku.edu.tw

摘要

我们提出了一种用于可扩展贝叶斯变量选择的对称随机扫描吉布斯采样器，通过即时计算所需量，无需存储完整的交叉乘积矩阵。基于边际相关性构建的数据驱动提议权重，会将抽样重点集中在有希望的候选变量上，同时，均匀混合成分可确保理论有效性。我们明确说明了如何根据信号与零相关性的比值选择调优参数，以确保充分探索后验分布。后验均值大小选择规则为中位概率模型提供了一种自适应替代方案，无需设定任意阈值，即可自动适应有效信号密度。在包含十万个预测变量的模拟中，该方法的灵敏度达到 1.000，精确率超过 0.76。将该方法应用于一项研究枯草芽孢杆菌核黄素生产的基因组数据集后，我们识别出六个基因，均得到先前采用其他方法开展的研究的验证。该方法的底层模型结合了狄拉克尖峰与宽尾先验以及拉普拉斯型收缩：狄拉克尖峰通过将非活动系数精确设为零来确保严格稀疏性，

而拉普拉斯型 slab 则通过局部—全局尺度混合机制，为活跃系数提供自适应正则化。

1 引言

尖峰与厚尾先验通过潜在指示变量对预测变量的纳入进行建模，为高维回归中的变量选择建立了贝叶斯框架（Mitchell and Beauchamp, 1988; George and McCulloch, 1993）。尽管这种先验在理论上具有优势，但在该先验下实际应用完全贝叶斯推断时，仍面临可扩展性问题（Ishwaran and Rao, 2005; Rocková, 2018; Rocková and George, 2014）。由于后验分布遍历的模型空间包含 2^p 种配置，用于探索该空间的马尔可夫链蒙特卡罗（MCMC）方法，其计算和内存需求会随着预测变量数量 p 的增加而变得难以承受。因此，人们普遍认为，基于精确 MCMC 的推断在超高维情形下不可行（Bhattacharya et al., 2016; Blei et al., 2017; Nishimura and Suchard, 2022; Rocková and George, 2018; Schrider and Kern, 2018; Scott and Berger, 2010; Shin et al., 2018; Zanella and Roberts, 2019）。

为应对这一实际限制，人们开发了以牺牲精确性换取可扩展性的替代策略。一种常见做法是先根据边际关联筛选变量，再对筛选后的子集应用贝叶斯方法，但这种临时筛选可能会扭曲后验推断（例如，Fan 和 Lv，2008年；Zhou 等人，2022年）。变分近似通过优化一个更简单的分布来提升计算速度，但会系统性地低估后验不确定性（例如，Carbonetto 和 Stephens，2012年；Wang 等人，2020年）。随机搜索方法（例如，Benner 等人，2016年；Gayevskiy 和 Goddard，2016年）通过限制模型空间来实现可处理性，但其后验分布必然以这些限制为条件。因此，推断会对算法调优参数十分敏感，例如最大模型规模和存储模型的数量；这些参数是固定的计算设置，而非由概率论根据

数据推断得到的随机变量。其他成功的方法则针对特定领域量身定制，例如基因组学；在这类领域中，方法会利用固有的低维相关结构，例如带状连锁不平衡矩阵。然而，这些专门方法是围绕基因组连锁不平衡矩阵的特定属性设计的，即其带状块对角结构。因此，它们无法推广到设计矩阵或相关结构不具备这些特征的其他回归问题（例如，Erbe 等人，2012年；Lloyd-Jones 等人，2019年）。这些有价值的策略都偏离了在完全贝叶斯尖峰与 slab 模型中进行精确推断这一目标。

本研究提出一种对称随机扫描吉布斯采样器（Roberts 和 Sahu，1997年；Levine 和 Casella，2006年），用于针对尖峰与厚尾先验下变量选择问题的精确后验分布，同时直接解决存储空间和每次迭代成本方面的障碍。我们不再通过系统扫描更新所有 p个坐标，而是只更新随机选取的一小部分 $m \ll p$ 个坐标。我们的方法采用对称选择方案，从同一个数据驱动分布中提议活跃变量和非活动变量，因此每次更新都是精确的吉布斯步骤，无需 Metropolis–Hastings 校正。随机扫描吉布斯采样器能够保证平稳分布始终为精确后验分布，而不受选择概率影响，前提是每个变量都有非零的选中概率。

这些设计选择在时间和内存方面都带来了显著的计算节省；参见第 ?? 节，了解与全扫描实现的简要比较。除了计算效率之外，该框架还具有更大的灵活性。数据驱动的提议权重将计算集中在最有希望的变量上，同时不会引入偏差，因为精确的吉布斯更新不受选择概率影响。我们使用通过解析积分消去回归系数后得到的折叠边际似然，这有助于改善马尔可夫链混合。对全局纳入概率采用层次 Beta-Bernoulli 先验，可使模型根据数据自适应地学习稀疏度。主要的权衡在于，与全扫描采样器相比，可能需要更多迭代才能充分探索模型

与全扫描采样器相比的空间。

本文其余部分安排如下。第 2 章介绍带有尖峰与厚尾先验的贝叶斯线性模型，推导折叠后验分布，并给出主要的抽样更新。第 3 章介绍实现细节，包括数据驱动扫描机制、即时计算策略和计算复杂度。第 4 章报告模拟结果。第 5 章对核黄素数据集进行实证研究。第 6 章以讨论作结。

2 方法

2.1 带拉普拉斯型厚尾先验的分层狄拉克尖峰

考虑如下线性回归模型:

$$\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n\right), \tag{1}$$

其中， $\boldsymbol{y} \in \mathbb{R}^n$ 为响应向量， $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ 为设计矩阵，包含 n 个观测值和p 个预测变量， $\boldsymbol{\beta} \in \mathbb{R}^p$ 为系数向量， σ^2 为误差方差， $\boldsymbol{I}_n$ 为 $n \times n$单位矩阵。

狄拉克尖峰与（拉普拉斯型）厚尾先验通过在 $\boldsymbol{\beta}$ 上采用带有二元纳入指标的尖峰与厚尾先验，实现稀疏性和收缩。

$$\boldsymbol{z} = (z_1, \ldots, z_p)^\top \in \{0,1\}^p,$$

其中， $z_j = 1$ 表示预测变量 j 被纳入模型。索引集 $\{1, \ldots, p\}$分为活动集 $A = \{j : z_j = 1\}$ ，其对应系数为 $\boldsymbol{\beta}_A$ ；以及非活跃集 $I = \{j : z_j = 0\}$ ，其对应系数为 $\boldsymbol{\beta}_I$。

在 Guan and Stephens（2011）的基础上结合局部-全局收缩（Carvalho 等人, 2010），我们采用 Park and Casella（2008）的尺度混合表示。层级先验

are

$$\boldsymbol{\beta}_A \mid \boldsymbol{\tau}_A^2, \kappa^2, \boldsymbol{z} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{D}_A\right), \ \ \boldsymbol{D}_A = \operatorname{diag}\left(\frac{\tau_j^2}{\kappa^2}\right)_{j \in A}, \tag{2}$$

$$\begin{aligned}
\boldsymbol{\beta}_I \mid \boldsymbol{z} &= \mathbf{0}, \\
\tau_j^2 \mid \lambda_1 &\overset{\text{iid}}{\sim} \operatorname{Exp}\left(\frac{\lambda_1^2}{2}\right), \ \ j = 1, \ldots, p, \\
\kappa^2 &\sim \operatorname{Gamma}\left(a_\kappa, b_\kappa\right), \\
\sigma^2 &\sim \operatorname{InvGamma}\left(a_\sigma, b_\sigma\right),
\end{aligned}$$

其中， $\boldsymbol{\tau}^2 = \left(\tau_1^2, \ldots, \tau_p^2\right)^\top$ 为局部尺度参数， κ^2 为全局收缩参数，$\lambda_1 > 0$ 控制总体稀疏性程度。对于稀疏性，我们使用

$$z_j \mid \pi \overset{\text{iid}}{\sim} \operatorname{Bern}\left(\pi\right), \ \ j = 1, \ldots, p, \tag{3}$$

$$\begin{aligned}
\pi \mid a_\pi, b_\pi &\sim \operatorname{Beta}\left(a_\pi, b_\pi\right), \\
a_\pi &\sim \operatorname{Gamma}\left(\alpha_a, \beta_a\right), \\
b_\pi &\sim \operatorname{Gamma}\left(\alpha_b, \beta_b\right).
\end{aligned} \tag{4}$$

超参数（α_a、 β_a、 α_b、 β_b）倾向于稀疏模型，同时允许数据确定有效的稀疏度。较小的速率参数 β_b on b_π 会促使b_π取较大的值，进而使 π 较小，因此模型也会更加稀疏。

For 对于每个系数，由（2）—（3）所隐含的条件先验可写为

$$\begin{aligned}
\beta_j \mid z_j, \tau_j^2, \kappa^2 &\sim \left(1 - z_j\right) \delta_0 + z_j \mathcal{N}\left(0, \frac{\tau_j^2}{\kappa^2}\right), \\
\tau_j^2 &\sim \operatorname{Exp}\left(\frac{\lambda_1^2}{2}\right),
\end{aligned} \tag{5}$$

因此，不活跃的 系数严格为零，而活跃系数服从高斯厚尾分布

其方差分解为 $\frac{\tau_j^2}{\kappa^2}$。对 τ_j^2 积分消去后，可得到 β_j 的拉普拉斯型全局-局部厚尾先验密度，将类似 ℓ_1 的厚尾特性与系数特异性自适应相结合。

在进行后验计算时，我们采用尺度混合增广。令局部精确率的倒数为 $\omega_j = \tau_j^{-2}$，则给定 $(\beta_j, \kappa^2, \lambda_1)$ 时，ω_j 的条件密度服从逆高斯分布：

$$\omega_j \mid \beta_j, \kappa^2, \lambda_1 \sim \mathrm{IG}\left(\mu_j, \lambda_1^2\right),$$

其中

$$\mu_j = \frac{\lambda_1}{|\beta_j|\,\kappa},$$

且 IG (μ, λ) 表示均值为 μ 、形状参数为 λ 的逆高斯分布 .

2.2 折叠边际似然

为避免重复进行 $p \times p$ 或 $n \times n$ 分解，我们利用共轭高斯结构对 $\boldsymbol{\beta}_A$ 进行解析积分，并遵循贝叶斯变量选择的折叠吉布斯策略（例如，Bhattacharya、Chakraborty 和 Mallick，2016；Guan 和 Stephens，2011）。边际似然为

$$p\left(\boldsymbol{y}|\boldsymbol{z}, \sigma^2, \kappa^2, \boldsymbol{\tau}^2\right) = \int p\left(\boldsymbol{y}|\boldsymbol{\beta}_A, \boldsymbol{z}, \sigma^2\right) p\left(\boldsymbol{\beta}_A|\boldsymbol{z}, \kappa^2, \boldsymbol{\tau}^2\right) d\boldsymbol{\beta}_A.$$

由此得到的折叠似然仅取决于 $(\boldsymbol{z}$、σ^2、κ^2、$\boldsymbol{\tau}^2)$，通过移除高度相关的系数来改善混合效果。我们仍从其完整条件分布中抽取 $\boldsymbol{\beta}_A$ ，以进行后验推断。

折叠似然可表示为关于 $n \times n$ 协方差矩阵的形式

$$\boldsymbol{S} = \sigma^2 \boldsymbol{I}_n + \boldsymbol{X}_A \boldsymbol{D}_A \boldsymbol{X}_A^\top,$$

因此

$$\log p\left(\boldsymbol{y}|\boldsymbol{z},\sigma^2,\kappa^2,\boldsymbol{\tau}^2\right)=-\frac{1}{2}\left(\log|\boldsymbol{S}|+\boldsymbol{y}^\top\boldsymbol{S}^{-1}\boldsymbol{y}\right)+\text{const.} \tag{6}$$

对式 (6) 进行朴素计算需要 $O(n^3)$ 次分解，但 $\boldsymbol{S}$ 的结构使得在 $|A|\ll n$ 时可以大幅减少计算量。

2.3 折叠似然的高效矩阵计算

为了利用矩阵行列式引理和伍德伯里恒等式，在不同活动集 A 之间实现高效计算，我们预先计算以下辅助量：

$$\boldsymbol{G}_0=\boldsymbol{X}^\top\boldsymbol{X}\in\mathbb{R}^{p\times p},$$

$$\boldsymbol{h}_0=\boldsymbol{X}^\top\boldsymbol{y}\in\mathbb{R}^p,$$

$$c_y=\boldsymbol{y}^\top\boldsymbol{y}\in\mathbb{R}.$$

这些量在初始化时计算一次；复杂度分析见章节?? 。

在马尔可夫链蒙特卡罗过程中，当 σ^2 发生变化时，我们只进行标量重缩放。具体而言，我们构造缩放后的长度为 p 的向量

$$\boldsymbol{h}=\frac{1}{\sigma^2}\boldsymbol{h}_0\in\mathbb{R}^p,$$

这样，对于任意活动集 A，$\boldsymbol{h}_A=\frac{\boldsymbol{h}_0}{\sigma^2}_A$ 都可立即得到。我们不构造完整的重缩放$p\times p$矩阵 $\frac{\boldsymbol{G}_0}{\sigma^2}$。相反，只要需要，我们就仅对活动格拉姆矩阵块进行重缩放，计算如下：

$$\boldsymbol{G}_{AA}=\frac{1}{\sigma^2}\boldsymbol{G}_{0,AA}.$$

根据式 (2) 中对 $\boldsymbol{D}_A$ 的定义，活跃集的精度矩阵为

$$\boldsymbol{M}=\boldsymbol{D}_A^{-1}+\frac{1}{\sigma^2}\boldsymbol{X}_A^\top\boldsymbol{X}_A=\operatorname{diag}\left(\frac{\kappa^2}{\tau_j^2}\right)_{j\in A}+\frac{1}{\sigma^2}\boldsymbol{G}_{0,AA}\in\mathbb{R}^{|A|\times|A|}. \tag{7}$$

矩阵 $\boldsymbol{M}$ 是活跃系数的精度矩阵，也是我们

计算策略的核心。

回忆一下 $\boldsymbol{S} = \sigma^2 \boldsymbol{I}_n + \boldsymbol{X}_A \boldsymbol{D}_A \boldsymbol{X}_A^\top$。伍德伯里恒等式已被用于实现快速贝叶斯变量选择（例如，Bhattacharya 等人，2016年；Johndrow、Orenstein 和 Bhattacharya，2020年），其内容如下：

$$\left(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V}^\top\right)^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C}^{-1} + \boldsymbol{V}^\top \boldsymbol{A}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{V}^\top \boldsymbol{A}^{-1},$$

适用于 $\boldsymbol{A} = \sigma^2 \boldsymbol{I}_n$、 $\boldsymbol{U} = \boldsymbol{X}_A$、 $\boldsymbol{C} = \boldsymbol{D}_A$ 和 $\boldsymbol{V} = \boldsymbol{X}_A$。代入这些取值可得

$$\begin{aligned} \boldsymbol{S}^{-1} &= \sigma^{-2}\boldsymbol{I}_n - \sigma^{-2}\boldsymbol{X}_A \left(\boldsymbol{D}_A^{-1} + \frac{1}{\sigma^2}\boldsymbol{X}_A^\top \boldsymbol{X}_A\right)^{-1} \boldsymbol{X}_A^\top \sigma^{-2} \\ &= \sigma^{-2}\boldsymbol{I}_n - \sigma^{-2}\boldsymbol{X}_A \boldsymbol{M}^{-1} \boldsymbol{X}_A^\top \sigma^{-2}, \end{aligned} \tag{8}$$

其中， $\boldsymbol{M}$ 定义见式（7）。将式（8）代入二次型可得

$$\begin{aligned} \boldsymbol{y}^\top \boldsymbol{S}^{-1}\boldsymbol{y} &= \sigma^{-2}\boldsymbol{y}^\top\boldsymbol{y} - \sigma^{-4}\boldsymbol{y}^\top \boldsymbol{X}_A \boldsymbol{M}^{-1}\boldsymbol{X}_A^\top \boldsymbol{y} \\ &= \sigma^{-2} c_y - \left(\frac{\boldsymbol{X}_A^\top \boldsymbol{y}}{\sigma^2}\right)^\top \boldsymbol{M}^{-1} \left(\frac{\boldsymbol{X}_A^\top \boldsymbol{y}}{\sigma^2}\right) \\ &= \sigma^{-2} c_y - \boldsymbol{h}_A^\top \boldsymbol{M}^{-1} \boldsymbol{h}_A, \end{aligned} \tag{9}$$

其中 $\boldsymbol{h}_A = \frac{\boldsymbol{h}_{0,\cdot\,A}}{\sigma^2} \in \mathbb{R}^{|A|}$。

矩阵行列式引理可提供类似的化简。对于维度相容的矩阵 $\boldsymbol{A}$、 $\boldsymbol{U}$、 $\boldsymbol{C}$、 $\boldsymbol{V}$，若 $\boldsymbol{A}$ 和 $\boldsymbol{C}$ 可逆，则有

$$\left|\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V}^\top\right| = \left|\boldsymbol{C}^{-1} + \boldsymbol{V}^\top \boldsymbol{A}^{-1}\boldsymbol{U}\right| \cdot |\boldsymbol{C}| \cdot |\boldsymbol{A}|,$$

如在标量情形中所用可行的尖峰与 slab 推断（例如，Carbonetto 和 Stephens，2012年；Biswas ，

Mackey 和 Meng，2022年）。应用于 $\boldsymbol{S} = \sigma^2 \boldsymbol{I}_n + \boldsymbol{X}_A \boldsymbol{D}_A \boldsymbol{X}_A^\top$ 可得到

$$\begin{aligned} |\boldsymbol{S}| &= \left| \boldsymbol{D}_A^{-1} + \boldsymbol{X}_A^\top \left(\sigma^2 \boldsymbol{I}_n\right)^{-1} \boldsymbol{X}_A \right| \cdot |\boldsymbol{D}_A| \cdot |\sigma^2 \boldsymbol{I}_n| \\ &= |\boldsymbol{M}| \cdot |\boldsymbol{D}_A| \cdot \left(\sigma^2\right)^n . \end{aligned}$$

取对数并使用 $|\boldsymbol{D}_A| = \prod_{j \in A} \binom{\tau_j^2}{\kappa^2}$，可得

$$\log |\boldsymbol{S}| = n \log \sigma^2 + \sum_{j \in A} \log \frac{\tau_j^2}{\kappa^2} + \log |\boldsymbol{M}|. \tag{10}$$

将 (9) 和 (10) 代入折叠对数似然，可得一个仅依赖于 $|A| \times |A|$ 矩阵 $\boldsymbol{M}$（通过 $\boldsymbol{M}^{-1}$、$\log |\boldsymbol{M}|$ 和二次型 $\boldsymbol{h}_A^\top \boldsymbol{M}^{-1} \boldsymbol{h}_A$）的表示。当 A 通过单次纳入或排除发生变化时，这种表示可实现高效更新；其复杂度将在第 ?? 章节中分析。

2.4 折叠式 Metropolis-within-Gibbs 采样器

2.4.1 逐坐标纳入移动

将折叠对数似然表示为关于 $\boldsymbol{M}$ 的函数，并预先计算 Gram 统计量后，我们通过逐坐标的 Metropolis-Hastings 移动更新 $\boldsymbol{z}$。在每次 MCMC 迭代中，我们按顺序扫描坐标 $j = 1, \ldots, p$。对于每个 j，我们尝试翻转z_j，提出添加移动（$z_j : 0 \to 1$）或删除移动（$z_j : 1 \to 0$）。接受概率取决于折叠对数似然 $\mathcal{L}(A)$ 的变化，以及 Beta-Bernoulli 模型所隐含的先验优势比。

令 $A = \{j : z_j = 1\}$ 表示当前活动集，并令 A' 表示通过翻转坐标 j得到的提议活动集。对数接受率可写为

$$\log r = \underbrace{\mathcal{L}(A') - \mathcal{L}(A)}_{\text{collapsed likelihood ratio}} + \underbrace{\log p\left(z_j' \mid \pi\right) - \log p\left(z_j \mid \pi\right)}_{\text{prior ratio}} + \underbrace{\log q\left(A \mid A'\right) - \log q\left(A' \mid A\right)}_{\text{proposal ratio}}.$$

由于提议是一次翻转，且反向移动也是同样的翻转，因此提议比率相互抵消，

$$\log q\left(A \mid A'\right) - \log q\left(A' \mid A\right) = 0.$$

因此，

$$\log r = \mathcal{L}\left(A'\right) - \mathcal{L}\left(A\right) + \log p\left(z_j' \mid \pi\right) - \log p\left(z_j \mid \pi\right).$$

给定 π，伯努利(π) 先验意味着

$$\log p\left(z_j' \mid \pi\right) - \log p\left(z_j \mid \pi\right) = \begin{cases} \log \pi - \log\left(1-\pi\right), & \text{add move } \left(z_j : 0 \to 1\right), \\ \log\left(1-\pi\right) - \log \pi, & \text{drop move } \left(z_j : 1 \to 0\right). \end{cases}$$

MH 接受概率为

$$\alpha = \min\left\{1, \exp\left(\log r\right)\right\},$$

通过抽取 $u \sim \text{Unif}\,(0,\ 1)$，并在 $\log\ u < \log\ r$ 时接受提议的翻转来实现。

每个逐坐标 MH 步骤都满足相对于边缘化目标分布的细致平衡，而顺序扫描则保持目标分布不变，并确保定期更新每个坐标。

主要的计算任务是高效地计算 $\mathcal{L}\,(A')\ - \mathcal{L}\,(A)$。我们通过乔列斯基分解初始化 $\boldsymbol{M}^{-1}$ 并计算一次 $|\boldsymbol{M}|$ 的对数，随后在接受添加/删除移动后，使用基于 Schur 补的扩边与降阶公式更新这些量（Ghosh 和 Clyde，2011）。为保证数值稳定性，我们跳过导致 Schur 补非正的提议。

2.4.2 秩一更新公式

我们推导了添加变量时的秩一更新公式。考虑添加变量 $j \notin A$ 以构成 $A' = A \cup \{j\}$。变量 j 的先验方差和精确率分别为 $d_j = \frac{\tau_j^2}{\kappa^2}$ 和 $d_j^{-1} = \frac{\kappa^2}{\tau_j^2}$。回顾 $\boldsymbol{G} = \frac{\boldsymbol{G}_0}{\sigma^2}$ 和 $\boldsymbol{h} = \frac{\boldsymbol{h}_0}{\sigma^2}$，因此

$$
\begin{aligned}
G_{jj} &= \frac{(G_0)_{jj}}{\sigma^2}, \\
\boldsymbol{g} = \boldsymbol{G}_{A,j} &= \frac{\boldsymbol{G}_{0,Aj}}{\sigma^2}, \\
h_j &= \frac{(h_0)_j}{\sigma^2}.
\end{aligned}
$$

令 $\boldsymbol{g} \in \mathbb{R}^{|A|}$ 表示 $\boldsymbol{G}$ 中对应于活跃变量与变量 j之间交互项的列。扩展后的精度矩阵可作如下分块表示

$$
\boldsymbol{M}' = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{g} \\ \boldsymbol{g}^\top & d_j^{-1} + G_{jj} \end{bmatrix} \in \mathbb{R}^{(|A|+1)\times(|A|+1)}.
$$

利用分块矩阵行列式恒等式，可得

$$
\log |\boldsymbol{M}'| = \log |\boldsymbol{M}| + \log s, \tag{11}
$$

其中

$$
s = d_j^{-1} + G_{jj} - \boldsymbol{g}^\top \boldsymbol{M}^{-1} \boldsymbol{g}. \tag{12}
$$

为保证数值有效性，舒尔补 s 必须为正。

为更新二次型，定义残差得分

$$
u = h_j - \boldsymbol{g}^\top \boldsymbol{M}^{-1} \boldsymbol{h}_A, \tag{13}
$$

调整当前条件后，衡量变量 j 与响应变量之间的关联。

活动集进行调整后，测量变量 $\boldsymbol{M}'$ 与响应变量之间的关联。 $\boldsymbol{M}'$ 的逆矩阵具有分块形式

$$(\boldsymbol{M}')^{-1} = \begin{bmatrix} \boldsymbol{M}^{-1} + s^{-1}\boldsymbol{M}^{-1}\boldsymbol{g}\boldsymbol{g}^{\top}\boldsymbol{M}^{-1} & -s^{-1}\boldsymbol{M}^{-1}\boldsymbol{g} \\ -s^{-1}\boldsymbol{g}^{\top}\boldsymbol{M}^{-1} & s^{-1} \end{bmatrix}. \tag{14}$$

令 $\boldsymbol{h}_{A'} = \left(\boldsymbol{h}_A^{\top}, h_j\right)^{\top} \in \mathbb{R}^{|A|+1}$。则

$$\begin{aligned} q(A') = \boldsymbol{h}_{A'}^{\top}(\boldsymbol{M}')^{-1}\boldsymbol{h}_{A'} &= \boldsymbol{h}_A^{\top}\boldsymbol{M}^{-1}\boldsymbol{h}_A + s^{-1}\left(h_j - \boldsymbol{g}^{\top}\boldsymbol{M}^{-1}\boldsymbol{h}_A\right)^2 \\ &= q(A) + \frac{u^2}{s}, \end{aligned} \tag{15}$$

其中 $q(A) = \boldsymbol{h}_A^{\top}\boldsymbol{M}^{-1}\boldsymbol{h}_A$，并且我们使用了（13）。

结合（11）和（15），并注意到

$$\log|\boldsymbol{D}_{A'}| - \log|\boldsymbol{D}_A| = \log d_j = \log\frac{\tau_j^2}{\kappa^2},$$

添加移动下折叠对数似然的变化为

$$\mathcal{L}(A') - \mathcal{L}(A) = -\frac{1}{2}\left[\log\frac{\tau_j^2}{\kappa^2} + \log s - \frac{u^2}{s}\right].$$

因此，添加移动的 MH 对数接受率为

$$\log r_{\text{add}} = -\frac{1}{2}\left(\log\frac{\tau_j^2}{\kappa^2} + \log s - \frac{u^2}{s}\right) + \log\frac{\pi}{1-\pi}.$$

如果接受该移动，我们使用（14）对 $\boldsymbol{M}^{-1}$ 进行增量更新。令 $\boldsymbol{t} = \boldsymbol{M}^{-1}\boldsymbol{g}$，则接受移动后的更新为

$$(\boldsymbol{M}')^{-1} = \begin{bmatrix} \boldsymbol{M}^{-1} + s^{-1}\boldsymbol{t}\boldsymbol{t}^{\top} & -s^{-1}\boldsymbol{t} \\ -s^{-1}\boldsymbol{t}^{\top} & s^{-1} \end{bmatrix}.$$

当 $A = \varnothing$ 时， $\boldsymbol{M}$ 为空泛，且 $\log |\boldsymbol{M}| = 0$。添加移动量简化为

$$
\begin{aligned}
s &= d_j^{-1} + G_{jj}, \\
u &= h_j, \\
q(A) &= 0, \\
q(A') &= \frac{h_j^2}{s},
\end{aligned}
$$

这是纳入第一个变量之前所采用的初始化。

对于满足 $A' = A \setminus \{j\}$ 且 $j \in A$ 的删除移动，可以利用 $\boldsymbol{M}^{-1}$ 的分块结构推导出类似的秩一更新。不失一般性，将索引重新排列，使 j 在 A 中位于最后，并作如下分块

$$
\boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{f} \\ \boldsymbol{f}^\top & g \end{bmatrix},
$$

其中 $g = s^{-1}$、 $\boldsymbol{E} \in \mathbb{R}^{(|A|-1)\times(|A|-1)}$、 $\boldsymbol{f} \in \mathbb{R}^{|A|-1}$ 和 $g \in \mathbb{R}$。舒尔补 s 与重新将变量 j 纳入 A' 时添加移动中的结果相同，其中 $g = s^{-1}$。这确保了添加移动和删除移动之间的一致性。行列式更新由分块行列式公式得出。由于 $|\boldsymbol{M}| = |\boldsymbol{M}/| s$，我们得到

$$
\log |\boldsymbol{M}'| - \log |\boldsymbol{M}| = -\log s.
$$

接下来，更新二次型。作如下分块

$$
\boldsymbol{h}_A = \begin{bmatrix} \boldsymbol{h}_- \\ h_m \end{bmatrix},
$$

其中 $\boldsymbol{h}_- \in \mathbb{R}^{|A|-1}$ 不包含索引 j，而 h_m 是与索引 j对应的分量。该

当前二次型可写为

$$q(A) = \boldsymbol{h}_-^\top \boldsymbol{E} \boldsymbol{h}_- + 2h_m \boldsymbol{f}^\top \boldsymbol{h}_- + g h_m^2.$$

根据分块矩阵求逆公式，$(|A|-1) \times (|A|-1)$ 顺序主子矩阵的逆为

$$(\boldsymbol{M}')^{-1} = \boldsymbol{E} - g^{-1} \boldsymbol{f}\boldsymbol{f}^\top,$$

删除 j 后，二次型变为

$$q(A') = \boldsymbol{h}_-^\top \left(\boldsymbol{E} - g^{-1} \boldsymbol{f}\boldsymbol{f}^\top \right) \boldsymbol{h}_- = \boldsymbol{h}_-^\top \boldsymbol{E} \boldsymbol{h}_- - g^{-1} \left(\boldsymbol{f}^\top \boldsymbol{h}_- \right)^2.$$

因此，二次型的变化为

$$q(A) - q(A') = 2h_m \boldsymbol{f}^\top \boldsymbol{h}_- + g h_m^2 + g^{-1} \left(\boldsymbol{f}^\top \boldsymbol{h}_- \right)^2.$$

对于折叠对数似然 $\mathcal{L}(A)$，移除变量 j时，只有涉及 $\frac{\tau_j^2}{\kappa^2}$、$\log |\boldsymbol{M}|$ 和 $q(A)$ 的项会发生变化。使用上述更新后，折叠对数似然的变化为

$$\mathcal{L}(A') - \mathcal{L}(A) = -\frac{1}{2} \left[-\log \frac{\tau_j^2}{\kappa^2} - \log s + 2h_m \boldsymbol{f}^\top \boldsymbol{h}_- + g h_m^2 + g^{-1} \left(\boldsymbol{f}^\top \boldsymbol{h}_- \right)^2 \right].$$

因此，删除移动的 MH 对数接受率为

$$\log r_{\text{drop}} = -\frac{1}{2} \left[-\log \frac{\tau_j^2}{\kappa^2} - \log s + 2h_m \boldsymbol{f}^\top \boldsymbol{h}_- + g h_m^2 + g^{-1} \left(\boldsymbol{f}^\top \boldsymbol{h}_- \right)^2 \right] + \log \frac{1-\pi}{\pi}.$$

2.4.3 完整采样周期

完整采样周期会更新其余参数。对于局部收缩参数τ_j^2，当 $j \in A$处于激活状态时，将高斯先验 $\beta_j \mid \tau_j^2,\ \kappa^2 \sim \mathcal{N}\left(0,\ \frac{\tau_j^2}{\kappa^2}\right)$与指数分布

混合先验 $\tau_j^2 \sim \text{Exp}\left(\frac{\lambda_1^2}{2}\right)$结合后，得到条件后验核（记作 $x = \tau_j^2$），

$$p\left(x \mid \beta_j, \kappa^2, \lambda_1\right) \propto x^{-1/2} \exp\left\{-\frac{1}{2}\left(\lambda_1^2 x + \frac{\kappa^2 \beta_j^2}{x}\right)\right\} \tag{16}$$

对于 $x > 0$。这是一个广义逆高斯（GIG）分布，其

$$\tau_j^2 \Big| \beta_j, \kappa^2, \lambda_1 \sim \text{GIG}\left(p = \tfrac{1}{2}, a = \lambda_1^2, b = \kappa^2 \beta_j^2\right).$$

对倒数 $\omega_j = \frac{1}{\tau_j^2}$ 进行处理，（16）意味着

$$p\left(\omega_j \mid \beta_j, \kappa^2, \lambda_1\right) \propto \omega_j^{-3/2} \exp\left[-\frac{1}{2}\left(\kappa^2 \beta_j^2 \omega_j + \frac{\lambda_1^2}{\omega_j}\right)\right]$$

对于 $\omega_j > 0$，这恰好属于采用标准参数化的逆高斯（IG）族

$$\omega_j \Big| \beta_j, \kappa^2, \lambda_1 \sim \text{IG}\left(\mu_j = \frac{\lambda_1}{|\beta_j| \kappa}, \lambda_j = \lambda_1^2\right).$$

局部方差为 $\tau_j^2 = \omega_j^{-1}$。逆高斯密度 $f(x|\mu, \lambda) \propto x^{-3/2} \text{exp}\left[-\frac{\lambda}{2\mu^2} x - \frac{\lambda}{2}\left(\frac{1}{x}\right)\right]$与 $\lambda = \lambda_1^2$ 和 $\mu = \frac{\lambda_1}{|\beta_j| \kappa}$ 相符。对于未激活的 $j \in I$，局部尺度从先验中抽取，

$$\tau_j^2 \Big| z_j = 0 \sim \text{Exp}\left(\frac{\lambda_1^2}{2}\right).$$

对于 κ^2，共轭性给出伽马后验。对于 $\kappa^2 \sim$ 伽马分布（a_κ，b_κ）以及 $\beta_j \mid \tau_j^2$，$\kappa^2 \sim \mathcal{N}\left(0, \frac{\tau_j^2}{\kappa^2}\right)$ 对于活跃的 $j \in A$，

$$\begin{aligned} p\left(\kappa^2 \mid \boldsymbol{\beta}_A, \boldsymbol{\tau}_A^2\right) &\propto \left(\kappa^2\right)^{a_\kappa - 1} \exp\left(-b_\kappa \kappa^2\right) \prod_{j \in A} \left(\kappa^2\right)^{1/2} \exp\left(-\frac{\kappa^2 \beta_j^2}{2\tau_j^2}\right) \\ &\propto \left(\kappa^2\right)^{a_\kappa + |A|/2 - 1} \exp\left[-\kappa^2 \left(b_\kappa + \frac{1}{2} \sum_{j \in A} \frac{\beta_j^2}{\tau_j^2}\right)\right]. \end{aligned}$$

因此，

$$\kappa^2 \mid \boldsymbol{\beta}_A, \boldsymbol{\tau}_A^2 \sim \text{Gamma}\left(a_\kappa + \frac{|A|}{2}, b_\kappa + \frac{1}{2}\sum_{j\in A}\frac{\beta_j^2}{\tau_j^2}\right).$$

当 $A = \emptyset$ 时，我们从先验伽马分布 (a_κ, b_κ) 中抽样。较大的 κ^2 会使活跃系数更强地全局收缩至零。

对于 σ^2，共轭性给出逆伽马后验分布。利用预先计算的 Gram 统计量，残差平方和为

$$\text{SSE} = c_y - 2\boldsymbol{\beta}_A^\top \boldsymbol{h}_{0,A} + \boldsymbol{\beta}_A^\top \boldsymbol{G}_{0,AA}\boldsymbol{\beta}_A,$$

从而避免显式计算 $\boldsymbol{X}\boldsymbol{\beta}$。在先验 $\sigma^2 \sim$ InvGamma (a_σ, b_σ) 下，条件后验分布为

$$\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{y} \sim \text{InvGamma}\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{\text{SSE}}{2}\right).$$

对于 $\boldsymbol{\beta}_A$，条件后验分布为高斯分布，其精度矩阵为

$$\boldsymbol{\Sigma}_A^{-1} = \frac{1}{\sigma^2}\boldsymbol{G}_{0,AA} + \text{diag}\left(\frac{\kappa^2}{\tau_j^2}\right)_{j\in A}. \tag{17}$$

We c重新计算一次乔列斯基分解 $\boldsymbol{\Sigma}_A^{-1} = \boldsymbol{R}^\top \boldsymbol{R}$。后验均值 is

$$\boldsymbol{m}_A = \boldsymbol{\Sigma}_A\left(\frac{1}{\sigma^2}\boldsymbol{h}_{0,A}\right),$$

其中 $\boldsymbol{h}_{0,A}$ 是 $\boldsymbol{h}_0$ 中与活动索引对应的子向量。随后通过求解包含 $\boldsymbol{R}$ 的三角方程组来抽取样本 $\boldsymbol{\beta}_A$ ，

$$\boldsymbol{\beta}_A = \boldsymbol{m}_A + \boldsymbol{R}^{-1}\boldsymbol{\epsilon},$$

其中 $\boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_{|A|}\right)$。

对于 π，层级 层次化 Beta-Bernoulli 结构采用超先验 $a_\pi \sim$ 伽马分布 (α_a, β_a) 和

$b_\pi \sim$ 伽马分布 (α_b, β_b)。先验均值 $\mathbb{E}(\pi \mid a_\pi, b_\pi) = \frac{a_\pi}{a_\pi + b_\pi}$ 表示活跃预测变量的预期占比。令 $K = \sum_j z_j$。若要使预期模型规模的中心位于 k，则设置 $b_\pi = a_\pi \left(\frac{p}{k} - 1\right)$。

给定当前的纳入向量 $\boldsymbol{z}$ 和超参数（a_π，b_π），π 的条件后验分布是共轭的，

$$\pi \mid \boldsymbol{z}, a_\pi, b_\pi \sim \mathrm{Beta}\left(a_\pi + |A|, b_\pi + (p - |A|)\right). \tag{18}$$

这会产生自动多重性校正（Scott 和 Berger，2010年）：随着 p 增加，先验会收缩 π，除非有充分证据表明存在许多活跃系数。通过 Metropolis–Hastings 探索（a_π，b_π）的联合条件后验分布。未归一化的后验密度为

$$p\left(a_\pi, b_\pi \mid \pi\right) \propto p\left(\pi \mid a_\pi, b_\pi\right) p\left(a_\pi\right) p\left(b_\pi\right), \tag{19}$$

其中

$$p\left(\pi \mid a_\pi, b_\pi\right) = \frac{\Gamma\left(a_\pi + b_\pi\right)}{\Gamma\left(a_\pi\right)\Gamma\left(b_\pi\right)} \pi^{a_\pi - 1}\left(1 - \pi\right)^{b_\pi - 1}, \tag{20}$$

$$p\left(a_\pi\right) = \mathrm{Gamma}\left(a_\pi | \alpha_a, \beta_a\right), \tag{21}$$

$$p\left(b_\pi\right) = \mathrm{Gamma}\left(b_\pi | \alpha_b, \beta_b\right). \tag{22}$$

我们在对数尺度上进行随机游走 Metropolis–Hastings。提议 $\log a_\pi^* = \log a_\pi + \epsilon_a$ 和 $\log b_\pi^* = \log b_\pi + \epsilon_b$，其中 ϵ_a，$\epsilon_b \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma^2_{\text{prop}}\right)$。对数接受率为

$$\begin{aligned}\log r = &\left[\log p\left(\pi \mid a_\pi^*, b_\pi^*\right) + \log p\left(a_\pi^*\right) + \log p\left(b_\pi^*\right) - \log a_\pi^* - \log b_\pi^*\right] \\ &- \left[\log p\left(\pi \mid a_\pi, b_\pi\right) + \log p\left(a_\pi\right) + \log p\left(b_\pi\right) - \log a_\pi - \log b_\pi\right],\end{aligned}$$

其中 $-\log a_\pi^{(*)} - \log b_\pi^{(*)}$ 项是雅可比行列式。以概率 $\min\{1, \exp(\log r)\}$ 接受。

此 Metropolis–Hastings 步骤仅涉及标量贝塔分布和伽马分布的密度，因此提高了稀疏先验的灵活性。层级先验不再固定（a_π，b_π）（即设定固定的先验期望模型规模），而是对（a_π，b_π，π，$\boldsymbol{z}$）设置层级先验，让数据决定先验的中心和形状

在 π 上的先验。进而，这会为纳入指标 $\boldsymbol{z}$ 生成后验分布，使其能够自动适应回归中的有效信号密度，同时保留贝塔-伯努利框架理想的多重性调整性质。

2.4.4 后验推断

该采样器生成后验抽样 $\left\{\left(\boldsymbol{z}^{(m)}, \boldsymbol{\beta}^{(m)}\right): \ m=1, \ldots\ldots, \ M\right\}$，来自 M 次迭代，其中 $z_j^{(m)} \in \{0, 1\}$ 表示预测变量 j 在第 m 次迭代中是否处于激活状态。后验纳入概率（PIP）为

$$\widehat{\mathrm{PIP}}_j = P\left(z_j = 1 \mid \boldsymbol{y}\right) \approx \frac{1}{M} \sum_{m=1}^{M} z_j^{(m)}, \tag{23}$$

对于 $j=1, \ldots\ldots, \ p$。向量 $\left(\widehat{\ \ }\mathrm{PIP}_1, \ldots\ldots, \ \widehat{\ \ }\mathrm{PIP}_p\right)$ 提供了一个数据驱动的预测变量排序，同时结合了关于 $\boldsymbol{z}$的先验。在贝叶斯变量选择中，PIP 的标准用法是将$\widehat{\ \ }\sum_j \mathrm{PIP}_j$ 解释为后验平均模型大小，并使用 PIP 对预测变量进行排序和选择（例如，Barbieri 和 Berger，2004年；Bogdan 等人，2011；Scott 和 Berger，2010年）。一种常见的选择规则是中位概率模型（Barbieri 和 Berger，2004年），该规则选择 PIP$\widehat{\ \ }$为$_j \geq 0.5$的变量。然而，当变量存在相关性时，许多变量的 PIP 可能略高于 0.5，从而导致纳入不理想的变量。此外，0.5 的截断值不会根据数据进行调整，因此忽略了后验不确定性。在信号较弱或预测变量相关性较高的情况下，即使是真实信号，其 PIP 也可能低于 0.5。中位数模型可能会将它们全部排除。

我们建议采用后验平均模型大小的 top-K 规则。令 $A = \{j : z_j = 1\}$ 表示给定模型中的活跃预测变量集合，$|A| = \mathbb{1}[z_j = 1] p \sum z_j$ 表示其大小。由于每个 $z_j \in \{0, 1\}$ 都是指示变量，因此 $\mathbb{E}\left[z_j \mid \boldsymbol{y}\right] = P\left(z_j = 1 \mid \boldsymbol{y}\right)$。根据期望的线性性质，

$$\mathbb{E}\left[|A| \mid \boldsymbol{y}\right] = \mathbb{E}\left[\sum_{j=1}^{p} z_j \,\middle|\, \boldsymbol{y}\right] = \sum_{j=1}^{p} \mathbb{E}\left[z_j \mid \boldsymbol{y}\right] = \sum_{j=1}^{p} P\left(z_j = 1 \mid \boldsymbol{y}\right),$$

因此，后验平均模型大小可以估计为

$$\hat{k} = \sum_{j=1}^{p} \widehat{\mathrm{PIP}}_j. \tag{24}$$

若要报告单个稀疏模型，请使用整数目标规模

$$k_\star = \max\left\{1, \min\left[p, \mathrm{round}\left(\hat{k}\right)\right]\right\}, \tag{25}$$

将四舍五入后的后验平均模型大小截断为 $[1, p]$。按后验纳入概率从大到小排序

$$\widehat{\mathrm{PIP}}_{(1)} \geq \widehat{\mathrm{PIP}}_{(2)} \geq \cdots \geq \widehat{\mathrm{PIP}}_{(p)},$$

并将阈值设为第 $k_\star$ 大的后验纳入概率，

$$\hat{t} = \widehat{\mathrm{PIP}}_{(k_\star)}. \tag{26}$$

所选活动集为

$$\widehat{A} = \left\{j : \widehat{\mathrm{PIP}}_j \geq \hat{t}\right\}. \tag{27}$$

若 $\hat{t}$ 处不存在并列值，则得到 $|\widehat{A}| = k_\star$。若出现并列值，$|\widehat{A}|$ 可能略大于 $k_\star$。最后，通过式（23）计算 $\widehat{\ }$后验包含概率j，根据式（24）计算 $\hat{k}$，使用式（25）–（26）确定 $k_\star$ 和 $\hat{t}$，$\widehat{\ }$并将式（27）中的 A 报告为所选预测变量集。

2.5 通过对称随机扫描实现可扩展计算

全扫描采样器在每次 MCMC 迭代中都会访问所有 p 坐标。对于超高维问题（Fan 和 Lv，2008），每次迭代的计算成本和存储需求会变得难以承受。为解决这些障碍，同时保持精确的后验分布目标，我们提出一种采用数据驱动的提议权重的对称随机扫描吉布斯采样器。选择权重 w_j 与状态无关，根据边际相关性计算一次后

续，并在整个 MCMC 过程中保持固定。这样便完全避免了使用 $p \times p$ 格拉姆矩阵，从而降低了每次迭代的计算成本。

2.5.1 即时计算格拉姆矩阵

为消除存储需求，我们仅预先计算长度为p 的量 es:

$$s_j = \boldsymbol{x}_j^\top \boldsymbol{y} = (\boldsymbol{h}_0)_j, \quad j = 1, \ldots, p, \tag{28}$$

$$t_j = \|\boldsymbol{x}_j\|^2 = (\boldsymbol{G}_0)_{jj}, \quad j = 1, \ldots, p, \tag{29}$$

$$c_y = \boldsymbol{y}^\top \boldsymbol{y}, \tag{30}$$

其中， $\boldsymbol{x}_j \in \mathbb{R}^n$ 是 $\boldsymbol{X}$的第 j列。

在马尔可夫链蒙特卡罗过程中，活动集与候选变量 j 之间的交叉乘积将即时计算：

$$\boldsymbol{G}_{A,j} = \boldsymbol{X}_A^\top \boldsymbol{x}_j \in \mathbb{R}^{|A|}. \tag{31}$$

活动集Gram矩阵$\boldsymbol{G}_{0,AA} \in \mathbb{R}^{|A|\times|A|}$ 将以增量方式维护。添加变量 j时，我们通过式（31）计算$\boldsymbol{G}_{A,j}$ 并扩展$\boldsymbol{G}_{0,AA}$；删除变量时，则删除对应的行和列。

2.5.2 数据驱动的提议权重

我们不扫描所有 p 坐标，而是在每次 $m \ll p$ 迭代中随机扫描一个子集。在超高维问题中，真实信号仅占极小比例，因此均匀选择效率低下。为了集中精力考察有希望的候选项，同时保证有效性，我们使用绝对边际相关性，

$$\rho_j = \frac{|s_j|}{\sqrt{t_j \cdot c_y}} = \frac{\left|\boldsymbol{x}_j^\top \boldsymbol{y}\right|}{\|\boldsymbol{x}_j\| \cdot \|\boldsymbol{y}\|} = \left|\operatorname{cor}(\boldsymbol{x}_j, \boldsymbol{y})\right|, \tag{32}$$

作为数据驱动权重。具有较高 ρ_j 的变量更有可能是活跃变量。

理想情况下，权重应与 PIPs $\pi_j = P(z_j = 1 \mid \boldsymbol{y})$ 成正比，但在运行马尔可夫链蒙特卡罗之前，这些值是未知的。借鉴确定性独立筛选（Fan 和 Lv，2008）的思想，我们使用边际相关性 ρ_j 作为未知后验纳入概率的代理。真正活跃的变量往往具有更高的边际相关性。这促使我们

$$w_j^{\text{data}} = \frac{\rho_j}{\sum_{\ell=1}^{p} \rho_\ell}.$$

然而，若对某些 j 有 $\rho_j = 0$，则 $w_j^{\text{data}} = 0$，这会违反遍历性。为确保对所有 j 都有 $w_j > 0$，我们构造一个防御性混合：

$$w_j = (1 - \epsilon) \cdot \frac{\rho_j}{\sum_{\ell=1}^{p} \rho_\ell} + \frac{\epsilon}{p}, \quad j = 1, \ldots, p, \tag{33}$$

其中 $\epsilon \in (0, 1)$（例如 $\epsilon = 0.1$）。数据引导分量将更新集中在有希望的变量上，而均匀分量则保证 $w_j \geq \frac{\epsilon}{p} > 0$，从而确保遍历性。归一化权重满足 $\sum_{j=1}^{p} w_j = 1$。

2.5.3 对称随机扫描吉布斯采样器

设 m 为每次迭代扫描的坐标数，其中 $m \ll p$。每次迭代，按照式（33）中的 $(w_1, \ldots, w_p)$，无放回地抽取 m 个不同索引 $\{j_1, \ldots, j_m\}$，取自 $\{1, \ldots, p\}$。然后，对每个选中的 j，在给定所有其他变量的条件下，对 z_j 执行精确的吉布斯更新。

其理念是以对称方式处理活跃变量和非活动变量。无论当前状态如何，我们都从相同的提议分布中进行采样。这对于获得正确的平稳分布至关重要。非对称随机扫描会根据某个变量当前处于活跃状态还是非活动状态，采用不同的选择概率。例如，可以优先选择活跃变量进行细化，或关注边际相关性较高的非活动变量，以便将其纳入。然而，这种依赖状态的选择

机制会引入一些复杂性。

首先，非对称选择需要在接受概率中进行提议比率校正。如果选择变量 j 的概率取决于是否选择了 $j \in A$，那么 MH 接受比率必须包含因子 $^{q_{(A|A')}\, q(A'|A)}$，其中 $q(\cdot|\cdot)$ 表示依赖状态的选择概率。计算这一校正会增加计算开销，并引入潜在的数值误差来源。

其次，理论分析会变得复杂得多。验证细致平衡需要仔细追踪活动集演变时选择概率如何变化；而要确保遍历性，则必须保证在正向和反向选择机制下，所有可能的转移都具有正概率。

第三，如果选择概率未得到适当平衡，非对称方案可能引入微妙的偏差。例如，如果活跃变量比非活跃变量更常被选中，那么即使后验分布更支持稀疏模型，马尔可夫链也可能难以向更稀疏的模型混合。

对称设计消除了这些复杂问题。由于选择概率 w_j仅取决于预先计算好的固定边际相关性 ρ_j，而不取决于当前状态 A，因此对于所有转移，提议比率 $^{q_{(A|A')}\, q(A'|A) = 1}$ 均为 1。结合直接从全条件分布中采样的精确吉布斯更新，对每个选定坐标的接受概率都为 1。由此得到的采样器在理论上简洁明了，能够精确保留后验分布，同时通过数据驱动权重将计算资源集中于有希望的变量。

对于每个选定的坐标 j，我们从其全条件分布中抽取 z_j：

$$P(z_j = 1 \mid z_{-j}, \boldsymbol{y}, \boldsymbol{\tau}^2, \kappa^2, \sigma^2, \pi) = \frac{1}{1 + \exp\left(-\ell_j\right)}, \tag{34}$$

其中， ℓ_j 表示比较纳入与排除变量 j 的对数优势比，

$$\ell_j = \log \frac{P\left(z_j = 1 \mid \cdot\right)}{P\left(z_j = 0 \mid \cdot\right)} = \mathcal{L}\left(A \cup \{j\}\right) - \mathcal{L}\left(A \setminus \{j\}\right) + \log \frac{\pi}{1-\pi}. \tag{35}$$

这里，$\mathcal{L}(\cdot)$ 表示折叠对数似然，而$\mathcal{L}\left(A \cup \{j\}\right) - \mathcal{L}(A \setminus \{j\})$ 使用前文推导的秩一更新公式计算。

若 $z_j = 0$ （当前未激活），则 $j \notin A$，因此 $A \setminus \{j\} = A$。添加 j 的对数优势为

$$\ell_j = -\frac{1}{2}\left(\log \frac{\tau_j^2}{\kappa^2} + \log s - \frac{u^2}{s}\right) + \log \frac{\pi}{1-\pi}, \tag{36}$$

其中 s 和 u定义见式 (12) 和 (13)。

若 $z_j = 1$ （当前已激活），则 $A \cup \{j\} = A$，且 $A \setminus \{j\}$ 为缩减后的活动集，因此我们计算保留 j 与将其移除的对数优势。利用删除移动公式，我们得到

$$\ell_j = -\frac{1}{2}\left(\log \frac{\tau_j^2}{\kappa^2} + \log s - \Delta q\right) + \log \frac{\pi}{1-\pi}, \tag{37}$$

其中 $\Delta q = 2h_m \boldsymbol{f}^\top \boldsymbol{h}_- + g h_m^2 + g^{-1}(\boldsymbol{f}^\top \boldsymbol{h}_-)^2$表示二次型的变化量。

给定 ℓ_j，我们抽样

$$z_j \sim \mathrm{Bern}\left(\frac{1}{1+\exp\left(-\ell_j\right)}\right). \tag{38}$$

如果采样值与当前状态不同，我们将相应地更新 A、 $\boldsymbol{G}_{0,\ AA}$、 $\boldsymbol{h}_A$、 $\boldsymbol{M}^{-1}$，并使用秩一公式记录 $|\boldsymbol{M}|$。

2.5.4 理论依据

对称随机扫描吉布斯采样器能够保持后验分布，因为：(i) 每次单坐标吉布斯更新都满足细致平衡；(ii) 不变核的凸组合仍是不变核（Liu 等，1995 年）；(iii) 均匀分量 $\frac{\epsilon}{p} > 0$确保了遍历性。关键在于，选择权重 $w_j = (1-\epsilon)\ \frac{\rho_j}{\sum \rho_\ell} + \frac{\epsilon}{p}$ 仅ℓ取决于预先计算的边际相关性 $\rho_j = |\mathrm{cor}(\boldsymbol{x}_j,\ \boldsymbol{y})|$，而不取决于当前状态 $\boldsymbol{z}$，因此无需

进行提议比率校正，并且每个被选中的坐标都通过精确的吉布斯抽样进行更新，接受概率为 1。

遍历性要求每个坐标的选择概率均为正，且全条件分布具有全支撑。由于对所有 j，$w_j \geq \frac{\epsilon}{p} > 0$，因此满足第一个条件。伯努利全条件分布对任意有限的对数优势 ℓ_j，都会以正概率赋值给 $z_j = 0$ 和 $z_j = 1$，因此满足第二个条件。任意两个纳入向量 $\boldsymbol{z}$，$\boldsymbol{z}' \in \{0, 1\}^p$ 均可通过单坐标翻转相互转换，且每次翻转的概率均为正，因此链具有不可约性。

3 实现

3.1 计算复杂度

我们总结全扫描采样器和对称随机扫描采样器的计算复杂度。辅助量 $\boldsymbol{G}_0 = \boldsymbol{X}^\top \boldsymbol{X}$、$\boldsymbol{h}_0 = \boldsymbol{X}^\top \boldsymbol{y}$ 和 $c_y = \boldsymbol{y}^\top \boldsymbol{y}$ 在初始化时计算一次。对于稠密的 $\boldsymbol{X}$，构造 $\boldsymbol{G}_0$ 的计算量为 $O(np^2)$，构造 $\boldsymbol{h}_0$ 的计算量为 $O(np)$；这些一次性开销会分摊到所有马尔可夫链蒙特卡罗迭代中。对于对称随机扫描采样器，不会构造完整的 $\boldsymbol{G}_0$ 矩阵；相反，只需预先计算 $O(p)$ 个对角元素 $t_j = \|\boldsymbol{x}_j\|^2$ 和边际统计量 $s_j = \boldsymbol{x}_j^\top \boldsymbol{y}$，其计算开销为 $O(np)$，存储空间为 $O(p)$。

全扫描采样器每次迭代都会访问所有 p 坐标，每次迭代的计算成本为$O(p|A|^2)$，且需要 $O(p^2)$ 存储空间来存储 $\boldsymbol{G}_0$。残差平方和计算为 $\mathrm{SSE} = c_y - 2\boldsymbol{\beta}_A^\top \boldsymbol{h}_{0,A} + \boldsymbol{\beta}_A^\top \boldsymbol{G}_{0,AA} \boldsymbol{\beta}_A$，计算成本为 $O(|A|^2)$，从而避免显式计算 $\boldsymbol{X}\boldsymbol{\beta}$。

对称随机扫描采样器的单次迭代成本较低。从提议分布中抽取 m 个坐标，需要 $O(m)$ 或 $O(p)$ 次运算，具体取决于抽样算法。对于每个 m 坐标，计算 $\boldsymbol{G}_{A,j} = \boldsymbol{X}_A^\top \boldsymbol{x}_j$ 需要 $O(n|A|)$ 次运算，计算 ℓ_j 需要 $O(|A|^2)$ 次运算，而当 z_j 发生变化时，更新 $\boldsymbol{M}^{-1}$ 需要 $O(|A|^2)$ 次运算。其余参数的吉布斯更新中，$\boldsymbol{\beta}_A$ 更新的乔列斯基分解需要 $O(|A|^3)$ 次运算，而 $\boldsymbol{\tau}^2$、κ^2、σ^2 和 π需要 $O(|A|)$ 次运算。单次迭代的总

成本为 $O(mn|A| + m|A|^2 + |A|^3)$。当 $|A| \ll n$ 和 $m \ll p$时，这比完整扫描的成本 $O(p|A|^2)$ 低得多。由于在稀疏问题中，$|A|$ 通常较小（例如，$|A| \leq 50$），因此 $|A|$ 中的三次项仍然易于处理。

采样器需要在 $O(np)$ 处存储数据矩阵 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$，在 s_1、.....、s_p($\boldsymbol{s} =$)$^\top$ 和 t_1、.....、t_p($\boldsymbol{t} =$)$^\top$ 处存储预计算向量，在$O(p)$处存储提议权重（w_1、.....、w_p），并在$O(|A|^2)$处存储活动集Gram矩阵 $\boldsymbol{G}_{0,\ AA}$ 以及精度矩阵的逆 $\boldsymbol{M}^{-1}$。关键在于，$p \times p$ 格拉姆矩阵 $\boldsymbol{G}_0$ 从未被构造或存储。

总之，每次迭代的计算成本从 $O(p|A|^2)$ 降至 $O(mn|A| + |A|^3)$，存储空间从 $O(p^2)$ 降至 $O(np)$。对于 $p = 10^5$ 和 $n = 500$，全扫描采样器需要约 80 GB 的存储空间来存储 $\boldsymbol{G}_0$，而对称随机扫描采样器只需 400 MB。

3.2 数值稳定性

我们的实现采用了数值保护措施，以确保稳健运行。在秩一更新中，我们要求满足舒尔补保护条件 $s > 10^{-12}$。违反此阈值（表明存在近共线性）的提议会被自动拒绝。

对于局部尺度，我们通过设置 $\beta_j^2 \leftarrow \max\left(\beta_j^2, 10^{-12}\right)$ 来防止 $|\beta_j|$ 过小，然后再计算 $\mu_j = \frac{\lambda_1}{|\beta_j|\kappa^2}$。遵循 Park and Casella (2008)、Makalic 和 Schmidt (2016) 以及 Li 和 Lin (2010) 的方法，我们使用 Michael- Schucany-Haas 算法从逆高斯分布中采样；如果采样器返回非有限值或非正值，则使用 $\omega_j \leftarrow \mu_j$ 作为备用方案。

对于 $\boldsymbol{\beta}_A$ 抽样，我们在每次迭代时都重新计算 $\boldsymbol{\Sigma}_A^{-1}$的乔列斯基分解，而不重复使用来自 $\boldsymbol{z}$ 更新的 $\boldsymbol{M}^{-1}$，从而避免重复秩一更新累积造成的舍入误差。

4 模拟

4.1 数据生成

我们从线性回归模型 $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 生成数据，其中 $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$。样本量为 $n = 500$，预测变量的数量为 $p \in \{10^4, 10^5\}$，满足 Fan 和 Lv（2008）的超高维判据，即对于 $\xi \in [0.36, 0.39]$，$\log(p) = O\left(n^{\xi}\right)$。噪声方差固定为 $\sigma^2 = 1$。真实系数向量 $\boldsymbol{\beta} \in \mathbb{R}^p$ 包含 $k = 10$ 个非零元素，

$$\beta_j = \begin{cases} 1, & j = 1, \ldots, 5, \\ -1, & j = 6, \ldots, 10, \\ 0, & j = 11, \ldots, p. \end{cases}$$

该配置产生了信噪比 SNR $\in [6.5, 8.5]$（Ročková 和 George，2018）和方差解释比例 PVE $\in [0.87, 0.89]$（Guan and Stephens，2011），表明真实信号是可识别的。在信号强度足够的情况下，挑战在于所提出的方法能否在该设置中识别真实预测变量。

设计矩阵 X 由块相关高斯分布生成（参见 Chaibub Neto 等人，2014；Foroughi Pour 和 Dalton，2018；Lu 和 Petkova，2014）。创建一个$p \times p$ 协方差矩阵 $\boldsymbol{\Sigma}$，其方差为 1，块大小为 20。对于同一块中满足 $j \neq k$ 的任意一对索引（j，k），我们令 $\Sigma_{jk} = \rho$，其中 $\rho \in$ {0.3, 0.7}，分别表示中等和强块内依赖；对于不同块中的索引，我们令 $\Sigma_{jk} = 0$。给定 $\boldsymbol{\Sigma}$，我们计算正则化 Cholesky 因子 C，并生成 X =ZC，其中 Z $\in \mathbb{R}^{n \times p}$ 的各元素均为独立同分布的 N (0, 1)。若最终抖动值为 jitter，则 X 的各行服从 N (0, $\boldsymbol{\Sigma}$ + jitter I_p)；当 Cholesky 分解在 jitter $= 0$ 时成功，则该分布简化为 N (0, $\boldsymbol{\Sigma}$)。为获得近奇异协方差矩阵的稳定 Cholesky 因子，我们采用自适应抖动策略。初始抖动值为 $= 0$，我们尝试计算 C = chol ($\boldsymbol{\Sigma}$ + jitter I_p)。若失败，则通过设定 jitter $= 10^{-8}$ 添加一个小的岭项；若再次失败，则每一步将 jitter 乘以

10，直至达到最大值 10^{-3}。这种指数回退策略能够以最小扰动 $\boldsymbol{\Sigma}$，成功完成 Cholesky 分解。

4.2 调优参数

均匀混合权重（ϵ）和每次迭代扫描的坐标数量（m）是对称随机扫描采样器中的调优参数。将这些参数与 p 进行适当缩放，对于可靠的推断至关重要。

均匀混合权重 ϵ 控制着效率与覆盖范围之间的平衡。较小的ϵ 会将更多概率集中于高信号变量，但可能对低信号候选变量探索不足。均匀分量 $\frac{\epsilon}{p}$ 无论其大小如何都能确保遍历性，因此较小的 ϵ 不会损害理论有效性。

m 需要重点考虑的是，必须充分频繁地访问真实信号变量，才能获得准确的后验纳入概率估计。对于边际相关性为 ρ_{j^*}的真实信号变量 j^*，每次迭代的选择概率为

$$w_{j^*} = (1-\epsilon)\cdot \frac{\rho_{j^*}}{\sum_{\ell=1}^{p}\rho_\ell} + \frac{\epsilon}{p}.$$

分母 $\sum\rho_\ell$ 主要由 $p-k$ 零变量主导。如果零变量具有典型的ℓ边际相关性，其阶数为 $\bar{\rho}_0$（源于抽样变异性），那么对于较大的 p， $\sum\rho_\ell \approx \bar{\rho}_0 \cdot p$ 为ℓ。因此， w_{j^*} 的数据引导分量按以下方式缩放：

$$\frac{\rho_{j^*}}{\sum_{\ell}\rho_\ell} \approx \frac{\rho_{j^*}}{\bar{\rho}_0 \cdot p} \propto \frac{1}{p}.$$

因此，即使采用数据引导的加权，真实信号的选择概率也会随着 p 增大而降低。

在整个马尔可夫链蒙特卡罗运行过程中，访问真实信号的预期次数为

$$V_{j^*} = n_{\text{iter}} \times m \times w_{j^*}.$$

为确保 PIP 估计可靠，我们要求 V_{j^*} 超过某个阈值（$V_{j^*} \gtrsim 1000$ 次访问）。由于 $w_{j^*} \propto \frac{1}{p}$，因此需要

$$n_{\text{iter}} \times m \propto p.$$

对于信号变量，其与 $\boldsymbol{y}$ 的总体边际相关可根据块相关设计计算得出。由于所有 10 个信号都位于第一个块中，且其系数之和为零（$\sum_{j=1}^{10}\beta_j = 0$），因此信号预测变量 x_j 与响应变量之间的协方差为

$$\operatorname{cov}(x_j, y) = \beta_j + \rho \sum_{k \neq j, k \leq 10} \beta_k = \beta_j + \rho\,(0 - \beta_j) = \beta_j\,(1 - \rho)\,.$$

y 的方差为 $\operatorname{var}(y) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} + \sigma^2 = 10(1-\rho) + 1 = 11 - 10\rho$，这是利用交叉项因系数均衡而相互抵消这一事实得到的。因此，信号边际相关为

$$\rho_{\text{signal}} = \frac{|1-\rho|}{\sqrt{11 - 10\rho}}.$$

对于 $\rho = 0.3$，我们有 $\rho_{\text{signal}} = \frac{0.7}{\sqrt{8}} = 0.247$。对于 $\rho = 0.7$，我们有 $\rho_{\text{signal}} = \frac{0.3}{\sqrt{4}} = 0.15$.

对于零变量，总体相关为零（即使是信号块内的零变量，平衡的信号系数也会相互抵消）。由于样本相关近似服从 $\mathcal{N}\left(0, \frac{1}{n}\right)$，因此 $|r|$ 服从均值为 $\sqrt{\frac{2}{\pi n}}$ 的半正态分布。当 $n = 500$ 时，样本相关绝对值的期望近似为

$$\bar{\rho}_{\text{null}} \approx \sqrt{\frac{2}{\pi n}} = \sqrt{\frac{2}{500\pi}} \approx 0.036.$$

相关性之和主要由 $p-k$ 零变量决定:

$$\sum_{\ell=1}^{p} \rho_\ell \approx k\rho_{\text{signal}} + (p-k)\,\bar{\rho}_{\text{null}} \approx p\,\bar{\rho}_{\text{null}}$$

对于 $p \gg k$。真实信号的选择概率变为

$$w_{j^*} = (1-\epsilon)\cdot\frac{\rho_{\text{signal}}}{p\bar{\rho}_{\text{null}}} + \frac{\epsilon}{p} = \frac{1}{p}\left[(1-\epsilon)\,\frac{\rho_{\text{signal}}}{\bar{\rho}_{\text{null}}} + \epsilon\right].$$

令 $R = \frac{\rho_{\text{signal}}}{\bar{\rho}_{\text{null}}}$表示信号与零假设之比。对于 $\rho = 0.3$,

$$R = \frac{0.247}{0.036} = 6.9.$$

对于 $\rho = 0.7$,

$$R = \frac{0.15}{0.036} = 4.2.$$

我们始终使用 $\epsilon = 0.1$。对于 $\rho = 0.3$ 且 $R = 6.9$, 系数为 $(1-\epsilon)\,R + \epsilon = 6.3$.对于 $\rho = 0.7$ 且 $R = 4.2$, 系数为 $(1-\epsilon)\,R + \epsilon = 3.9$. 为确保 $V_{j^*} \geq 1000$ 次访问, 我们要求 $n_{\text{iter}} \times m \geq \frac{1000p}{c}$, 其中 $c \in \{6.3, 3.9\}$ 取决于 ρ。

对于 $p = 10^4$, 我们使用 $m = 500$ 和 $n_{\text{iter}} = 10000$, 得到 $n_{\text{iter}} \times m = 5 \times 10^6$. $\rho = 0.3$ 的期望访问次数为

$$V_{j^*} = 5 \times 10^6 \times \frac{6.3}{10000} = 3150.$$

$\rho = 0.7$ 的期望访问次数为

$$V_{j^*} = 5 \times 10^6 \times \frac{3.9}{10000} = 1950.$$

对于 $p = 10^5$, w 使用 $m = 1000$ 和 $n_{\text{iter}} = 30000$, 得到 $n_{\text{iter}} \times m = 3 \times 10^7$. Th e

$\rho = 0.3$ 的期望访问次数为

$$V_{j^*} = 3 \times 10^7 \times \frac{6.3}{100000} = 1890.$$

$\rho = 0.7$ 的期望访问次数为

$$V_{j^*} = 3 \times 10^7 \times \frac{3.9}{100000} = 1170.$$

所有设置均确保 $V_{j^*} > 1000$ ，从而可靠地进行PIP估计。

4.3 初始值

为在避免从空模型开始的同时促进稀疏性，我们按如下方式生成稀疏初始值。将目标模型规模设为 $k_{\text{target}} = 20$ ，并从所有大小为 $\min(k_{\text{target}}, p)$ 的活动集中均匀随机抽取一个初始活动集$A^{(0)} \subset \{1,\ldots,p\}$。随后，对 $j \in A^{(0)}$设定 $z_j^{(0)} = 1$，对 $j \notin A^{(0)}$设定 $z_j^{(0)} = 0$，并将相应的回归系数初始化为 $j \in A^{(0)}$时取 $\beta_j^{(0)} \sim \mathcal{N}(0, 0.1^2)$，否则取 $\beta_j^{(0)} = 0$。局部尺度通过 $\tau^2{}_{(0)\,j} \sim \text{Exp}(\text{rate} = 3\lambda_1^2)$初始化，以施加强初始收缩。选择超参数 $\left(\pi^{(0)}, a_\pi^{(0)}, b_\pi^{(0)}\right)$，使先验期望模型规模与 k_{target}相匹配。我们设定 $a_\pi^{(0)} = 1$、$b_\pi^{(0)} = a_\pi^{(0)}\left(\frac{p}{k_{\text{target}}} - 1\right)$ 和 $\pi^{(0)} = \frac{k_{\text{target}}}{p}$，使初始贝塔先验下 π 的期望为$\mathbb{E}\left[\pi^{(0)}\right] = \frac{k_{\text{target}}}{p}$，因此 $\mathbb{E}(|A|) \approx k_{\text{target}}$。最后，将其余全局参数初始化为 $\kappa^{2\,(0)} = 1$ 和 $\sigma^{2\,(0)} = 1$。

4.4 结果

表 1 汇总了 10 次独立重复实验中的变量选择表现。我们比较了两种选择规则：k 规则，即选择后验纳入概率（PIP）最高的 k 个变量，其中 k 根据模型规模的后验分布估计；以及中位数模型，即纳入所有 PIP 大于 0.5 的变量。

表 1：变量选择性能

p	ρ	阈值	TP	FP	灵敏度	精确率
10^4	0.3	$\hat{k}$-规则	10.0 (0.0)	2.8 (0.6)	1.000 (0.000)	0.783 (0.039)
		中位数	10.0 (0.0)	0.2 (0.4)	1.000 (0.000)	0.982 (0.038)
	0.7	$\hat{k}$-规则	10.0 (0.0)	2.5 (1.0)	1.000 (0.000)	0.804 (0.055)
		中位数	10.0 (0.0)	0.1 (0.3)	1.000 (0.000)	0.991 (0.029)
10^5	0.3	$\hat{k}$-规则	10.0 (0.0)	3.1 (0.3)	1.000 (0.000)	0.764 (0.017)
		中位数	10.0 (0.0)	0.0 (0.0)	1.000 (0.000)	1.000 (0.000)
	0.7	$\hat{k}$-规则	10.0 (0.0)	3.1 (0.6)	1.000 (0.000)	0.765 (0.033)
		中位数	10.0 (0.0)	0.1 (0.3)	1.000 (0.000)	0.991 (0.029)

数值为均值，括号内为标准差。TP = 真正例；FP = 假阳性；FN =假阴性。

所提出的方法在所有模拟设置下均表现出良好的变量选择性能。两种选择规则的灵敏度均达到 1.000，在每次重复实验中都正确识别出全部 10 个真实预测变量。这表明，采用狄拉克尖峰和拉普拉斯型宽松先验的折叠吉布斯采样器能够充分恢复真实的稀疏信号，即使在具有强块内相关性的超高维情形（$p = 10^5$）（$\rho = 0.7$）下也不例外。

中位数模型在所有设置下均具有很高的精确率（≥ 0.982）。对于 $p = 10^5$（ $\rho = 0.3$），中位数模型在全部 10 次重复实验中均实现完全恢复（TP $= 10$，FP $= 0$，精确率 $= 1$），这表明后验纳入概率高度集中于真实模型。即使在强相关（$\rho = 0.7$）的情况下，中位数模型平均产生的假阳性也至多为 0.1–0.2，对应的精确率高于 0.98。

相比之下，$\hat{k}$ 规则倾向于将模型规模高估 2–3 个变量（$\hat{k} \approx 12.5$– 13.1，而真实模型规模为 $k = 10$），导致精确率较低（$0.76 - 0.80$）。这反映了 k 规则的保守性：无论后验纳入概率（PIP）如何，它都会选取排名靠前的 k 个变量，因此纳入低于 0.5 阈值的边际预测变量。中位数模型采用明确的后验纳入概率阈值，可自然地防止此类假发现。

结果表明，该方法具有稳健性。维度增加十倍，从 $p = 10^4$ 到$p = 10^5$ ，并未降低选择准确率；事实上，维度较高时，中位数模型的表现略好，这与关于稀疏

高维回归中后验集中的理论结果一致。同样，块内相关性从 $\rho = 0.3$ 增加到$\rho = 0.7$，对性能的影响也很小，尽管相关设计通常会因同一块内预测变量之间的多重共线性而给变量选择带来挑战。

对于 $p = 10^4$，采样器每次重复约需 3.5 分钟，扩展至 $p = 10^5$ 时约需 60–75 分钟。从 $p = 10^4$ 增至 $p = 10^5$（对于 p 增幅为 $10\times$，约为 17–$22\times$）的耗时增长，既反映了更大的坐标扫描规模（$m = 1000$ 对比 $m = 500$），也反映了为确保对真实信号有足够访问次数而需要运行更长的链（$n_{\text{iter}} = 30000$ 对比 $n_{\text{iter}} = 10000$）。尽管如此，该方法在计算上仍然可行：使用 $p = 10^5$ 个预测变量运行完整的马尔可夫链蒙特卡罗（MCMC）约需一小时，且只需单核，因此适用于现代高维应用。在 $p = 10^5$ 时变异性更高（标准差为 12–23 分钟），这可能是由于访问到的活跃变量数量有所变化，以及抽样期间相关矩阵运算所致。

5 实证研究

5.1 数据与实现

我们使用从 `hdi` R 包获得的核黄素生产数据集评估该方法。这是一个高维基因组学基准数据集，包含 $n = 71$ 个观测值，记录了枯草芽孢杆菌 中对数转换后的核黄素产量，以及 $p = 4088$ 个基因表达预测变量（$\frac{p}{n} \approx 57.6$）。

设计矩阵和响应变量均已中心化并标准化。我们计算边际关联得分 $\rho_j = \frac{|\boldsymbol{X}_j^\top \boldsymbol{y}|}{\|\boldsymbol{X}_j\| \|\boldsymbol{y}\|}$。将排名前 $k_0 = 20$ 的得分与其余得分进行比较，得到信号与零假设的比率 $R = 4.11$。给定探索参数 $\epsilon = 0.1$，有效浓度为$c = (1 - \epsilon)\, R + \epsilon = 3.80$。为确保信号的期望访问次数为 >1000，我们要求 $m \geq \frac{1000 \cdot c}{p \cdot n_{\text{iter}}}$。我们设定 $n_{\text{iter}} = 60000$, $m = 350$（对应 19,500 次信号期望访问），预热期为 10000 次，并每隔 2 次抽取一次。

采样器随机初始化。均匀抽取大小为 $k_{\text{target}} = 20$ 的活动集 $A^{(0)}$，其中 $\beta_j^{(0)} \sim \mathcal{N}(0, 0.1^2)$ 用于 $j \in A^{(0)}$。设置超参数以引入稀疏性：$a_\pi^{(0)} = 1$, $b_\pi^{(0)} = \frac{p}{k_{\text{target}}} - 1 = 203.4$, 得到 $\mathbb{E}\left[\pi^{(0)}\right] \approx 0.0049$。

5.2 结果与比较

使用 $\hat{k}$ 规则（选择排名前 $\hat{k} = \sum_j$ 个后验包含概率j 变量），我们得到 $\hat{k} = 5.85$，并选出六个基因（表2）。误差方差估计值为 $\hat{\sigma}^2 = 0.195$。排名第一的基因 YOAB_at（后验包含概率=0.870）呈现强负关联，并且在先前的主要研究中均被识别出来。文献中也经常选出基因 YXLD_at 和 YXLE_at。我们的方法通过纳入 YXLF_at，找回了完整的 yxl 操纵子三联体，并为 ARGF_at 和 CARB_at 赋予非零概率；这两个基因较少出现在以往的分析中。

所选基因形成了具有生物学解释性的簇。三个基因 YXLD_at、YXLE_at 和 YXLF_at属于 yxl 操纵子，该操纵子编码膜相关调控蛋白。共同选中这三个基因，凸显了该方法捕捉操纵子层面信号的能力。基因 ARGF_at 和 CARB_at 参与精氨酸和嘧啶生物合成，表明它们可能通过氮代谢与核黄素生产存在代谢联系。

表2：核黄素数据集的变量选择结果

Gene	PIP	$\hat{\beta}$（标准差）	先前研究
YOAB_at	0.870	−0.417 （0.179）	a、b、c、d、e、f
ARGF_at	0.409	−0.156 (0.194)	c
YXLE_at	0.378	−0.181 (0.241)	b、c、e、f
YXLD_at	0.296	−0.140 (0.222)	a、b、c、d、e、f
YXLF_at	0.155	−0.064 (0.154)	e
CARB_at	0.130	−0.047 (0.124)	d

[a]Bühlmann等（2014年），[b]Javanmard与Montanari（2014年），[c]Chichignoud等（2016年），[d]Bar等（2020年），[e]Guo等（2021年），[f]Gong等（2021年）。

6 讨论

本文通过对带有拉普拉斯型收缩的狄拉克尖峰与宽尾先验采用对称随机扫描吉布斯采样器，构建了一个适用于超高维回归的可扩展贝叶斯变量选择框架。我们将狄拉克尖峰先验、局部-全局收缩和折叠吉布斯采样整合到统一框架中。尽管这些组成部分各自在文献中均有出现，但将它们结合起来，便得到一个具有互补特性的采样器。我们利用伍德伯里恒等式、格拉姆矩阵预计算和矩阵行列式引理来降低计算复杂度。

对称随机扫描吉布斯采样器通过避免完整的坐标扫描和存储 $p \times p$ 格拉姆矩阵，将可扩展性拓展至超高维情形。其关键思想是，基于预先计算的边际相关性 $\rho_j = |\mathrm{cor}(\boldsymbol{x}_j, \boldsymbol{y})|$构造的数据驱动的提议权重，会将抽样工作集中在有希望的候选项上，同时均匀混合成分可确保遍历性。由于这些权重仅取决于固定量，而非当前状态，因此无需进行提议比率校正，且采样器能够保持精确的后验分布。

第4章节中的调参分析为选择m 和 n_{iter}提供了一个有原则的框架。公式 $m \geq {}^{1000}{\cdot c}^{\ p}{\cdot}_{n_{\mathrm{iter}}}$，其中 $c = (1-\epsilon)\,R + \epsilon$ 取决于信号与零假设之比 R，可确保真实信号获得足够的访问次数，从而实现可靠的PIP估计。这种对探索-利用权衡的明确刻画，为高维马尔可夫链蒙特卡罗中的调参提供了有原则的指导。

ˆ我们提出 k-规则，作为中位概率模型的替代方案，用于报告ˆ ⌒一组选定的预测变量。该规则选取PIP最高的 $k = \sum_j$变量ˆ，其中 k 为后验平均模型大小。这种方法具有若干实用特性：它完全由后验分布决定，无需外部阈值；能够适应数据中的有效信号密度；当许多变量的后验纳入概率集中在某个任意截断值附近时，它还能提供一种自然的概括方式。

ˆ 模拟结果显示， k 规则的灵敏度达到 1.000，但平均会纳入 2–3 个

额外变量，体现出其更具包容性的特点。在遗漏真实信号代价高昂的应用中，$\hat{}\, k$ 规则可能更受青睐。

对核黄素数据集的应用为该方法提供了外部验证。 k-规则选出的全部六个基因，都曾在至少一项采用其他高维推断方法的研究中被识别出来（表2），这为所选信号及其生物学可解释性提供了外部支持。

所提出的贝叶斯变量选择方法采用对称随机扫描采样器，通过子抽样而非分布近似实现可扩展性，在降低计算成本的同时保留吉布斯采样器的理论保证。此外，用于后验包含概率的 k-规则在模拟和经验研究中均表现良好。

参考文献

[Bar 等人（2020）]

Bar, H. Y., Booth, J. G., & Wells, M. T.（2020）。一种用于广义线性模型变量选择的可扩展经验贝叶斯方法。《计算与图形统计学杂志》，*29*（3），535–546。

[Barbieri 和 Berger（2004年）]

Barbieri, M. M., & Berger, J. O. (2004年). 最优预测模型选择。《统计年鉴》，*32*(3), 870–897。

[Benner 等人（2016）]

Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., & Pirinen, M. (2016). FINEMAP：利用全基因组关联研究的汇总数据进行高效变量选择。《生物信息学》，*32*(10), 1493–1501。[Bhattacharya 等人（2016）]

Bhattacharya, A.、Chakraborty, A. 和 Mallick, B.K.（2016）。高维回归中使用高斯尺度混合先验的快速抽样。《生物统计学》，103(4)，985–991。

[Biswas 等（2022）] Biswas, N., Mackey, L., & Meng, X.-L. (2022). 可扩展尖峰与板块先验。arXiv预印本 arXiv:2204.01668。

[Blei等(2017)] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). 变分推断：统计学家综述。美国统计学会会刊, *112*(518), 859–877。

[Bogdan 等人(2011)] Bogdan, M., Ghosh, J. K., & Tokdar, S. T. (2011). Benjamini-Hochberg程序与一些用于多重检验的贝叶斯规则的比较。载于 J. M. Bernardo 等人编, 贝叶斯统计学*9* （第 211–230 页）。牛津大学出版社。

[B{ü}hlmann 等人(2014)] Bühlmann, P., Kalisch, M., & Meier, L. (2014). 面向生物学应用的高维统计学。《统计学及其应用年度评论》, *1*, 255–278.

[Carbonetto & Stephens(2012)] Carbonetto, P., & Stephens, M. (2012). 用于回归中贝叶斯变量选择的可扩展变分推断及其在遗传关联研究中的准确性。《贝叶斯分析》, *7*(1), 73–108.

[Carvalho 等人(2010)] Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). 用于稀疏信号的马蹄形估计量。《生物统计学》, *97*(2), 465–480.

[Chaibub Neto 等人(2014)] Chaibub Neto, E., Bare, J. C., & Margolin, A. A. (2014). 将模拟研究作为设计实验：在“大 p、小 n”情形下对惩罚回归模型的比较。*PLOS ONE*, *9*(10), e107957。

[Chichignoud 等人(2016)] Chichignoud, M., Lederer, J., & Wainwright, M. J. (2016). 一种实用方案和快速算法，可在提供最优性保证的前提下调整 Lasso。机器学习研究杂志，*17*(231)，1–20。

[Erbe 等人(2012)] Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A., & Goddard, M. E. (2012). 利用填补的高密度单核苷酸多态性面板提高奶牛品种内及品种间基因组预测的准确性。乳业科学杂志, *95*(7), 4114–4129.

[Fan & Lv（2008）] Fan, J., & Lv, J. (2008). 针对超高维特征空间的确定性独立筛选。《皇家统计学会会刊：*B*辑》, *70*(5), 849–911.

[Foroughi pour 和 Dalton(2018)] Foroughi pour, A., & Dalton, L. A. (2018). 贝叶斯模型下具有分块对角协方差结构的特征选择启发式算法。*BMC*生物信息学，*19*(Suppl 3), 70.

[Gayevskiy & Goddard（2016）] Gayevskiy, V., & Goddard, M. E. (2016). 数千个连锁基因座对表型的影响的同时估计。遗传学，*203*(4), 1599–1610.

[George & McCulloch（1993）] George, E. I., & McCulloch, R. E. (1993). 通过吉布斯抽样进行变量选择。美国统计学会会刊，*88*(423), 881–889.

[Ghosh & Clyde（2011）] Ghosh, J.,& Clyde, M.A.（2011）。线性

和二元回归。美国统计学会会刊，*106*(495)，1041–1052。

[Gong 等人（2021）] Gong, S., Zhang, K., & Liu, Y.（2021）。带高维成对筛选的惩罚线性回归。《中华统计学志》，*31*，391–420。

[Guan & Stephens（2011）] Guan, Y., & Stephens, M.（2011）。用于全基因组关联研究及其他大规模问题的贝叶斯变量选择回归。《应用统计年鉴》，*5*（3），1780–1815。

[Guo 等人（2021）] Guo, Z., Renaux, C., Bühlmann, P., & Cai, T.T.（2021）。高维组推断及其在层级检验中的应用。《电子统计学期刊》，*15*（2），6633–6676。

[Ishwaran 与 Rao（2005）] Ishwaran, H., & Rao, J. S.（2005）。尖峰与厚尾变量选择：频率学派与贝叶斯策略。《统计年鉴》，*33*(2)，730–773。

[Javanmard 与Montanari（2014）] Javanmard, A., & Montanari, A.（2014）。高维回归的置信区间与假设检验。机器学习研究杂志，*15*，2869–2909。

[Johndrow 等（2020）] Johndrow, J. E., Orenstein, P., & Bhattacharya, A. (2020). 用于马蹄先验的可扩展近似马尔可夫链蒙特卡罗算法先验。 机器学习研究杂志，*21*（73），1–61。

[Levine & Casella（2006年）] Levine, R. A., & Casella, G.（2006年）。优化随机扫描吉布斯采样器。 多元分析杂志，*97*（10），2071–2100。

[Li & Lin（2010年）] Li, Q., & Lin, N. (2010). 贝叶斯弹性网络。《贝叶斯分析》, *5*(1), 151–170。

[Liu 等人（1995）] Liu, J. S., Wong, W. H., & Kong, A. (1995). 各种扫描方式下吉布斯采样器的协方差结构与收敛速度。《皇家统计学会会刊：*B*辑》, *57*(1), 157–169。

[Lloyd-Jones et al.(2019)] Lloyd-Jones, L.R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K. E., Wang, H., Zheng, Z., Magi, R., Esko, T., et al. (2019). 基于汇总统计量进行贝叶斯多元回归可改善多基因预测。《*Nature Commu- nications*》, *10*(1), 5086.[Lu & Petkova(2014)]

Lu, F., & Petkova, E. (2014). 心理筛查工具开发背景下变量选择方法的比较研究。《医学统计学》,*33*(3), 401–421.

[Makalic & Schmidt(2016)] Makalic, E., & Schmidt, D. F.(2016). 一种简单的马蹄形估计量采样器。《*IEEE*信号处理快报》,*23*(1), 179–182.

[Michael 等人(1976)] Michael, J. R., Schucany, W. R., & Haas, R. W.(1976). 使用多根变换生成随机变量。《美国统计学家》, *30*(2), 88–90.

[Mitchell 和 Beauchamp（1988）] Mitchell, T. J., & Beauchamp, J. J.（1988）。线性回归中的贝叶斯变量选择。美国统计学会会刊, *83*(404), 1023–1032。

[Nishimura 与 Suchard（2022）] Nishimura, A., & Suchard, M. A.（2022）。用于加速“large n, large p”贝叶斯稀疏回归吉布斯抽样的先验预条件共轭梯度方法。美国统计学会会刊，*117*(540), 2059–2072。

[Park 与 Casella（2008）] Park, T. 与 Casella, G.（2008）。贝叶斯套索。美国统计学会会刊，*103*(482), 681–686。

[Roberts 与 Sahu（1997年）] Roberts, G. O. 与 Sahu, S. K.（1997年）。Gibbs采样器的更新方案、相关结构、分块和参数化。《皇家统计学会会刊：*B*辑》，*59*(2), 291–317。

[Rockov{á}（2018）] Rocková, V. (2018). 使用连续尖峰与厚尾先验对稀疏信号进行贝叶斯估计。统计年鉴，*46*(1), 401–437。

[Rockov{á} 与 George(2014)] Rocková, V., & George, E.I. (2014). EMVS：贝叶斯变量选择的 EM 方法。美国统计学会会刊，*109*(506), 828–846。

[Rockov{á} 与 George(2018)] Rocková, V., & George, E. I. (2018). 尖峰与厚尾 LASSO。美国统计学会会刊，*113*(521), 431–444。

[Schrider 与 Kern(2018)] Schrider, D. R., & Kern, A. D.（2018）。群体遗传学的监督式机器学习：一种新范式。遗传学趋势，*34*(4), 301–312。

[Scott和Berger（2010年）] Scott, J. G., & Berger, J. O.（2010）。变量选择问题中的贝叶斯与经验贝叶斯多重性调整。《统计年鉴》，*38*(5)，2587–2619。

[Shin等（2018）] Shin, M., Bhattacharya, A., & Johnson, V.E.（2018）。在超高维情形下使用非局部先验密度进行可扩展的贝叶斯变量选择。《中华统计学志》，*28*（2），1053–1078。

[Wang等（2020）] Wang, Y., Blei, D. M., & Cunningham, J. P.（2020）。后验塌缩与潜变量不可识别性。神经信息处理系统进展，*33*，5443–5455。

[Zanella和Roberts（2019）] Zanella, G. 与 Roberts, G.（2019）。可扩展的重要性退火与贝叶斯变量选择。《皇家统计学会会刊：*B*辑》，*81*(3)，489–517。

[Zhou 等人（2022）] Zhou, Q.、Guan, Y.、Carbonetto, P. 与 Stephens, M.（2022）。通过贝叶斯回归和连续收缩先验进行多基因预测。《自然·通讯》，*13*(1)，4569。