arXiv:2303.05898v1 [stat.ME] 10 Mar 2023

用于高维马蹄回归的协变量数据学习

Claudio Busatto

佛罗伦萨大学统计学、计算机科学与“G. Parenti”应用系，意大利佛罗伦萨

Mark van de Wiel

阿姆斯特丹大学医学中心流行病学与数据科学系、阿姆斯特丹公共卫生研究所，荷兰阿姆斯特丹

摘要

高维数据常见于临床基因组学研究，用于推断与特定性状相关的预测变量。提升预测性能的一种方法是纳入基于先验知识或既往研究得出的预测变量信息。这类信息也称为“协变量数据”。为此，我们开发了一种新颖的贝叶斯模型，将协变量数据纳入高维回归框架，称为信息性马蹄回归（infHS）。该方法将回归参数的先验方差对协变量变量进行回归，从而改善变量选择和预测。我们实现了吉布斯采样器和变分近似算法。前者适用于中等维度的应用，除预测外也可用于后验推断；后者则凭借其计算效率，能够处理数量极大的变量。我们通过模拟研究展示了纳入协变量数据的优势。最后，我们证明，在两个基因组学应用中，infHS 的表现优于其他竞争方法。

关键词：贝叶斯推断、变分贝叶斯、信息性收缩先验、马蹄先验、协变量信息。

1 引言

高维数据集的分析是许多科学领域的主要研究方向。尤其是由于预算限制，临床研究常常需要处理样本量相对较少、但数据量庞大的数据，例如基因表达或全基因组甲基化水平数据。

或实际条件的限制。我们采用回归模型，利用这类数据研究临床结局。由于参数数量 p 远超样本量，n我们提出一种方法，将（先验）外部知识纳入估计过程，以提升模型的整体表现。这类外部信息来源称为协变量数据（补充数据；Neuenschwander 等人，2016），因为它为协变量提供了额外信息。我们考虑两种不同类型：连续型数据，例如 p以往研究中的值；或类别型数据，例如所属群组，如染色体或特定基因组区域。

有几种方法可以在回归框架中纳入一种辅助信息来源（Tai and Pan，2007；Boonstra 等人，2013）。一种流行的方法是（稀疏）组LASSO（Yuan and Lin，2006年；Simon 等人，2013），它使用一个共同的超参数对所有组中的变量进行惩罚。当协变量组的数量较多时，这种方法颇具吸引力，但缺乏灵活性，在其他情形下也无法进行局部自适应，因而会导致次优结果（M¨unch 等人，2019年）。较新的研究着重于估计自适应惩罚，并为每个组设定特定的先验方差（Van de Wiel 等人，2019年；Velten and Huber，2019；M¨unch 等人（2019年））。然而，这些方法的适用范围有限，因为它们要么只处理一种（离散）协变量数据源，要么只适用于一种特定类型的结局。相比之下，Van Nee 等人（2021年）提出了一种岭回归方法，通过将局部方差对协变量数据进行回归，从而支持多个协变量数据源。利用他们的方法，可以针对每个协变量数据源分别估计辅助数据回归参数，最终再使用一个权重向量将其合并，其中每个权重对应一个协变量数据源的重要性。

在此，我们提出一种适用于线性回归和二元回归、能够纳入多种协变量信息的新型贝叶斯方法。具体而言，我们提出马蹄回归（Carvalho 等人，2010年）的一种推广，称为信息性马蹄回归模型（infHS）。我们遵循Van Nee 等人（2021年）的研究，将局部方差对协变量进行回归。与Kpogbezan 等人（2019年）不同，后者仅将马蹄先验用于单一的双组协变量数据源，而我们的模型能够灵活处理不同类型的协变量数据，既适用于连续型协变量数据预测变量，也适用于离散型协变量数据预测变量。此外，该模型还可通过 Probit 回归扩展到二元结局（Albert 和 Chib，1993年）。与Van Nee 等人（2021年）不同，我们的方法能够处理稀疏情形，并联合估计与不同协变量数据源相关的协变量回归参数，从而避免针对每个协变量数据源分别进行多次回归。

我们首先提出一种吉布斯采样器，用于迭代更新后验参数。我们引入一种新颖的拒绝采样方法，用于从局部方差的非解析全条件分布中进行采样。随着变量数量增加，我们采用Bhattacharya 等人（2016）提出的计算方法，从多元正态密度中高效采样。为使该方法适用于特别大型的 p 场景，我们借鉴M¨unch 等人（2019年）的技术，开发了联合后验分布的变分贝叶斯近似，以高效优化变分分布的目标密度，从而得到一种计算时间随 p线性增长的算法。通过模拟和两个数据应用，我们

证明了纳入协变量信息有助于提升预测和变量选择的效果；后者在马蹄先验设定下尤为重要。

本文结构安排如下。第2 节介绍模型的层级结构并讨论参数化方式。第3节和第4节分别介绍吉布斯抽样算法以及联合后验分布的变分贝叶斯近似。第5 节通过模拟研究说明协变量信息对变量选择的益处；第6 节则展示了该模型在两个数据集上的应用，一个来自遗传学，另一个来自癌症基因组学。最后，我们在第7节进行讨论并提出可能的扩展。

2 模型

令$\mathbf{y} \in \mathbb{R}^n$ 为响应向量，令$\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$ 为设计矩阵，其第一列为全 $\mathbf{1}$。我们使用广义线性模型（GLM），以$\mathbf{y}$ 为因变量、以$\mathbf{X}$ 为自变量进行回归，回归系数向量为 $\boldsymbol{\beta} = [\beta_0\ \beta_1 \ldots \beta_p]^\intercal$。

$$
\begin{aligned}
Y_i \mid \mathbf{x}_i, \boldsymbol{\beta} &\overset{ind}{\sim} p\left(Y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) \\
E_{Y_i|\mathbf{x}_i,\boldsymbol{\beta}}\left(Y_i\right) &= g^{-1}\left(\mathbf{x}_i^\intercal \boldsymbol{\beta}\right), \quad i = 1, \ldots, n,
\end{aligned}
$$

其中，$\mathbf{x}_i = [1\ x_{i1} \ldots x_{ip}]^\intercal$表示与观测值 i相关的协变量集合。

沿用Carvalho 等人 (2010年)的做法，马蹄先验会将回归参数局部收缩至零附近，并为 $\boldsymbol{\beta}$提供稀疏解。我们假设每个 β_j服从正态先验分布，其方差分解为全局尺度参数 τ 和局部收缩参数 λ_j。形式上，

$$
\begin{aligned}
\beta_0 \mid \sigma^2, \tau, \lambda_0 &\sim \mathcal{N}\left(0, \sigma^2\tau^2\lambda_0^2\right), \\
\lambda_0 &\sim \mathcal{C}^+(0, 1), \\
\beta_j \mid \sigma^2, \tau, \lambda_j &\sim \mathcal{N}\left(0, \sigma^2\tau^2\lambda_j^2\right), \quad j = 1, \ldots, p.
\end{aligned}
$$

假设有 D 个不同的协变量数据源$\mathbf{Z}_d \in \mathbb{R}^{p\times m_d}$ 可用，其中 $\sum_d m_d = M$ 和 $d = 1, \ldots, D$。为了捕捉外部信息的影响，Van Nee 等人（2021年）引入参数 $\omega_d > 0$ 和 $\boldsymbol{\gamma}_d \in \mathbb{R}^{m_d}$, $d = 1, \ldots, D$，将协变量特异性收缩 λ_j, $j = 1, \ldots, p$建模为$\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_D]$的函数。参数 $\boldsymbol{\gamma}_d$表示与矩阵$\mathbf{Z}_d$ 相关的回归系数向量，并针对每个组 d分别估计；而协变量数据权重 $\omega_d > 0$ 则用于建模组 d 的相对重要性，并被引入以整合不同的协变量数据源。这里，参数 $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_1 \ldots \boldsymbol{\gamma}_D\right]$ 通过从其全条件分布中抽样来联合更新。这样便能自然地整合协变量数据源，并可将分组权重 ω_d 排除在模型之外。因此，我们模型中的先验分布层次结构

为

$$\begin{aligned}
\lambda_j \mid \mathbf{Z}, \boldsymbol{\gamma} &\sim \mathcal{C}\left(\sum_{d=1}^{D}\left(\mathbf{z}_j^d\right)^{\intercal}\boldsymbol{\gamma}_d, s_0^2\right)\cdot \mathbb{I}_{(\lambda_j>0)}, \quad j=1,\ldots,p,\\
\boldsymbol{\gamma}_d \mid \boldsymbol{\Sigma}_{\gamma_d} &\sim \mathcal{N}_{m_d}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma_d}\right), \quad d=1,\ldots,D,\\
\tau &\sim \mathcal{C}^{+}\left(0,1\right),\\
\sigma^2 &\sim \mathcal{IG}\left(v,q\right),
\end{aligned} \tag{1}$$

其中 $\mathcal{C}$ 表示柯西分布， $\mathcal{C}^{+}$ 表示其正半部分。这里，我们对先验局部方差的位置参数进行建模，主要是因为这样可以从多元正态分布中联合抽样协变量数据系数 $\boldsymbol{\gamma}$ （见第 3 节）。请注意，当协变量数据不具信息性且对每个 $\boldsymbol{\gamma}_d \to 0 \quad d$时，该模型便简化为 Carvalho 等人 (2010年)提出的普通马蹄先验，其中 $s_0^2 = 1$。

在 Van de Wiel 等人 (2019年)中，作者使用经验贝叶斯估计量分别针对每个来源，根据数据估计先验协方差矩阵 $\boldsymbol{\Sigma}_d$。然而，这种方法计算负担较重，因为它需要运行多个 MCMC 链直至收敛。这里我们考虑组特异尺度参数κ_d^2 和岭型先验 $\boldsymbol{\Sigma}_d = \kappa_d^2 \mathbf{I}_{m_d}$。这种方法的主要优势在于计算效率高，因为只需更新先验尺度参数 κ_d^2。对于 κ_d^2，其共轭先验分布为

$$\kappa_d^2 \sim \mathcal{IG}\left(a_d, b_d\right). \tag{2}$$

参数 κ_d^2 在模型深处发挥作用，可以认为它们对整体估计过程的影响较小。因此，对 a_d 和 b_d 作非信息性选择便已足够。

3 后验推断

本节介绍一种吉布斯采样器，通过从参数的全条件分布中采样，迭代更新参数。我们在以下假设下，详细说明线性回归模型的算法

$$y_i = \mathbf{x}_i^{\intercal}\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right), \quad i=1,\ldots,n.$$

然而，P´olya-Gamma 潜变量（Polson 等人，2013）或 Probit 广义线性模型（Albert 和 Chib，1993年）可用于扩充模型，从而使 $\boldsymbol{\beta}$ 二元回归模型中的全条件分布为高斯分布。

1. 对 $\boldsymbol{\beta}$ 和 σ^2进行采样。 $\boldsymbol{\beta}$ 和 σ^2 的全条件分布为

$$\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}, \sigma^2, \tau^2, \lambda_0^2, \boldsymbol{\lambda} \sim \mathcal{N}_{p+1}\left(\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{X}^{\intercal}\mathbf{y}, \sigma^2\boldsymbol{\Sigma}_{\beta}^{\star}\right), \quad \boldsymbol{\Sigma}_{\beta}^{\star} = \left(\mathbf{X}^{\intercal}\mathbf{X} + \tau^{-2}\boldsymbol{\Lambda}^{-2}\right)^{-1}, \tag{3}$$

$$\sigma^2 \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \tau^2, \lambda_0^2, \boldsymbol{\lambda} \sim \mathcal{IG}\left(v + \frac{n+p+1}{2}, q + \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \frac{\beta_0^2}{2\tau^2\lambda_0^2} + \frac{1}{2\tau^2}\sum_{j=1}^{p}\frac{\beta_j^2}{\lambda_j^2}\right),$$

其中 $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_p)$。从（3）中采样需要求逆 $p \times p$协方差矩阵 $\mathbf{\Sigma}_{\beta}^{\star}$；当协变量的数量 p 增加时，这在计算上将变得不可行，因为朴素求逆的复杂度为 $\mathcal{O}\left(p^3\right)$。对于高维问题，我们采用 Bhattacharya 等人(2016) 和 Johndrow 等人(2020) 提出的方法，以降低对回归参数 $\boldsymbol{\beta}$进行采样所需的 $\mathcal{O}\left(n^2 p\right)$ 运算量。

2. 抽样 $\boldsymbol{\lambda}$、 $\boldsymbol{\gamma}$ 和 $\boldsymbol{\kappa}^2$。对于 λ_j、 $j \neq 0$和 $\boldsymbol{\gamma}$ 在（1）中的先验分布并非共轭。为此，我们采用Geweke（1993年）提出的数据增广步骤，以构建共轭框架，并通过从多元正态分布中抽样来联合更新参数 $\boldsymbol{\gamma}$。

命题 1. 设 x 和 y 为满足以下条件的随机变量

$$x \sim \mathcal{C}\left(m, s^2\right) \quad \text{and} \quad y \sim \mathtt{t}_v\left(m, s^2\right),$$

其中， t_v 表示 Student-t 分布。若 $v = 1$，则 $X \sim Y$。

命题 2. 设 x 和 θ 为满足以下条件的随机变量

$$x \mid m, s^2, \varphi^2 \sim \mathcal{N}\left(m, s^2\varphi^2\right) \quad \text{and} \quad \varphi^2 \sim \mathcal{IG}\left(\frac{v}{2}, \frac{v}{2}\right),$$

则 $x \mid m$、 s^2、 $v \sim \mathtt{t}_v\left(m\right.$、 $\left.s^2\right)$（Geweke，1993年）。

根据命题 1 和 2，引入潜因子 $\varphi_j^2 \sim \mathcal{IG}$（1/2、1/2）后， λ_j 的先验分布可方便地重写为截断于 0 的正态分布。这样，协变量数据回归系数 $\boldsymbol{\gamma}$ 的正态先验便具有共轭性，且参数可以通过从多元正态分布中抽样轻松更新。参数 λ_j、 φ_j^2、 $\boldsymbol{\gamma}_d$ 的先验分布组可重写为

$$\begin{aligned}
\lambda_j \mid \mathbf{Z}, \boldsymbol{\gamma}, \varphi_j^2 &\sim \mathcal{N}\left(\sum_{d=1}^{D}\left(\mathbf{z}_j^d\right)^{\mathsf{T}} \boldsymbol{\gamma}_d, s_0^2\varphi_j^2\right) \cdot \mathbb{I}_{(\lambda_j > 0)}, \\
\varphi_j^2 &\sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{2}\right), \quad j = 1, \ldots, p, \\
\boldsymbol{\gamma}_d \mid \kappa_d^2 &\sim \mathcal{N}_{m_d}\left(\mathbf{0}, \kappa_d^2 \mathbf{I}_{m_d}\right), \\
\kappa_d^2 &\sim \mathcal{IG}\left(a_d, b_d\right), \quad d = 1, \ldots, D.
\end{aligned}$$

令 $\mu_j = \sum_{d=1}^{D}\left(\mathbf{z}_j^d\right)$ $\boldsymbol{\gamma}_d$, $\boldsymbol{\lambda} = [\lambda_1 \ldots \lambda_p]$ 和 $_{\mathsf{T}}$ $\mathbf{\Phi}^2 =$ 对角矩阵 $(\boldsymbol{\varphi})$，其中 $\boldsymbol{\varphi} = \left[\varphi_1^2, \ldots, \varphi_p^2\right]$。$_{\mathsf{T}}$

增广模型中 λ_j、 $\boldsymbol{\gamma}$、 φ_j^2 和 κ_d^2 的全条件分布为

$$\begin{aligned}
\pi\left(\lambda_j \mid \mathbf{Z}, \beta_j, \sigma^2, \tau^2, \boldsymbol{\gamma}, \varphi_j^2\right) &\propto \lambda_j^{-1} e^{-\frac{\beta_j^2}{2\sigma^2\tau^2\lambda_j^2} - \frac{\lambda_j^2}{2s_0^2\varphi_j^2} + \frac{\mu_j\lambda_j}{s_0^2\varphi_j^2}} \cdot \mathbb{I}_{(\lambda_j>0)}, \\
\varphi_j^2 \mid \mathbf{Z}, \lambda_j, \boldsymbol{\gamma} &\sim \mathcal{IG}\left(1, \frac{1}{2} + \frac{(\lambda_j - \mu_j)^2}{2s_0^2}\right), \\
\boldsymbol{\gamma} \mid \mathbf{Z}, \boldsymbol{\lambda}, \boldsymbol{\varphi}, \boldsymbol{\kappa}^2 &\sim \mathcal{N}_M\left(\boldsymbol{\Sigma}_\gamma^\star \left(\mathbf{Z}^\intercal \boldsymbol{\Phi}^{-2} \boldsymbol{\lambda}\right), s_0^2 \boldsymbol{\Sigma}_\gamma^\star\right), \\
\kappa_d^2 \mid \boldsymbol{\gamma}_d &\sim \mathcal{IG}\left(a_d + \frac{m_d}{2}, b_d + \frac{\boldsymbol{\gamma}_d^\intercal \boldsymbol{\gamma}_d}{2}\right),
\end{aligned} \tag{4}$$

其中 $\boldsymbol{\Sigma}_\gamma^\star = \left(\mathbf{Z}\ \boldsymbol{\Phi}^{-2}\mathbf{Z} + s_0^2\mathbf{D}_\kappa^{-1}\right)^{-1}$和 $\mathbf{D}_\kappa$ =对角矩阵 $\left(\kappa_1^2\mathbf{1}_{m_1}、\ \ldots\ldots、\ \kappa_D^2\mathbf{1}_{m_D}\right)$。关于如何在不计算未知归一化常数的情况下对参数进行抽样的详细信息，见补充材料第 B.1节。

所提出的框架具有计算优势：局部方差λ_j^2 可以并行计算，从而有望提高模型效率。

3. 抽样 τ^2。 在线性回归模型中，针对全局尺度参数 τ 的半柯西先验与 β_j，$j = 0, \ldots\ldots, p$的先验方差不共轭，且模型具有高斯误差。因此，我们采用 Makalic 和 Schmidt （2016）提出的策略来更新 τ。作者指出，半柯西分布可以表示为逆伽马分布的尺度混合，从而能够对 τ^2进行共轭更新。

命题 3. 如果 $\tau^2 \mid \zeta \sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\zeta}\right)$和 $\zeta \sim \mathcal{IG}\left(\frac{1}{2},\ 1\right)$，那么 $\tau \sim \mathcal{C}^+(0,\ 1)$。

根据命题 3 ， τ^2 和 ζ 的全条件分布可写成闭式形式。具体而言，参数通过从以下密度中抽样来更新：

$$\begin{aligned}
\tau^2 \mid \boldsymbol{\beta}, \sigma^2, \lambda_0^2, \boldsymbol{\lambda}, \zeta &\sim \mathcal{IG}\left(\frac{p}{2} + 1, \frac{1}{\zeta} + \frac{\beta_0^2}{2\sigma^2\lambda_0^2} + \frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\beta_j^2}{\lambda_j^2}\right), \\
\zeta \mid \tau^2 &\sim \mathcal{IG}\left(1, 1 + \frac{1}{\tau^2}\right).
\end{aligned}$$

补充材料的 A.1 节讨论了如何使用与 λ_0^2相似的命题 β_0来更新截距的局部尺度参数3。

4 变分贝叶斯近似

当协变量数量极大时，第2 节介绍的方法在计算上不可行。本节将讨论联合后验分布的一种高效近似。变分近似的推导请参阅补充材料的第C.1 节。

在第2节介绍的模型假设下，均值场近似可得

$$q(\boldsymbol{\theta}) = q(\boldsymbol{\beta}) \cdot q\left(\lambda_0^2\right) \cdot q(\lambda_1) \cdot \ldots \cdot q(\lambda_p) \cdot q(\psi_0) \cdot q\left(\varphi_1^2\right) \cdot \ldots \cdot q\left(\varphi_p^2\right) \cdot q(\boldsymbol{\gamma}) \cdot q\left(\kappa_1^2\right) \cdot \ldots \cdot q\left(\kappa_D^2\right) \cdot q\left(\tau^2\right) \cdot q(\zeta) \cdot q\left(\sigma^2\right),$$

其中， $q(\boldsymbol{\beta})$ 和 $q(\boldsymbol{\gamma})$ 分别表示 $\beta_0, \ldots, \beta_p$ 和 $\gamma_1, \ldots, \gamma_M$ 的联合变分分布。在算法的每次迭代中，变分因子按如下方式更新:

$$\begin{aligned}
q^\star(\boldsymbol{\beta}) &= \mathcal{N}_{p+1}\left(\boldsymbol{\mu}_\beta^\star, \mathbb{E}_{\sigma^2}\left[\sigma^2\right] \boldsymbol{\Sigma}_\beta^\star\right), \\
&\quad \boldsymbol{\mu}_\beta^\star = \boldsymbol{\Sigma}_\beta^\star \mathbf{X}^\intercal \mathbf{y}, \quad \boldsymbol{\Sigma}_\beta^\star = \left(\mathbf{X}^\intercal \mathbf{X} + \mathbb{E}_{\lambda_0^2 \cdot \boldsymbol{\lambda} \cdot \tau^2}\left[\tau^{-2} \boldsymbol{\Lambda}^{-2}\right]\right)^{-1}, \\
q^\star\left(\lambda_0^2\right) &= \mathcal{IG}\left(1, a_0^\star\right), \\
&\quad a_0^\star = \mathbb{E}_\zeta\left[\psi_0^{-1}\right] + \frac{1}{2} \mathbb{E}_{\beta_0 \cdot \sigma^2 \cdot \tau^2}\left[\frac{\beta_0^2}{\sigma^2 \tau^2}\right], \\
q^\star(\psi_0) &= \mathcal{IG}\left(1, k_0^\star\right), \\
&\quad k_0^\star = 1 + \mathbb{E}_{\lambda_0^2}\left[\lambda_0^{-2}\right], \\
q^\star(\lambda_j) &\propto \lambda_j^{-1} \exp\left\{-\frac{a_j^\star}{\lambda_j^2} - b_j^\star \lambda_j^2 + c_j^\star \lambda_j\right\} \cdot \mathbb{I}_{(\lambda_j > 0)}, \quad j = 1, \ldots, p, \\
&\quad a_j^\star = \frac{1}{2} \mathbb{E}_{\beta \cdot \sigma^2 \cdot \tau^2}\left[\frac{\beta_j^2}{\sigma^2 \tau^2}\right], \quad b_j^\star = \frac{1}{2 s_0^2} \mathbb{E}_{\varphi^2}\left[\varphi_j^{-2}\right], \quad c_j^\star = \frac{1}{s_0^2} \mathbf{z}_j^\intercal \mathbb{E}_{\boldsymbol{\gamma} \cdot \varphi^2}\left[\frac{\boldsymbol{\gamma}}{\varphi_j^2}\right], \\
q^\star\left(\varphi_j^2\right) &= \mathcal{IG}\left(1, d_j^\star\right), \\
&\quad d_j^\star = \frac{1}{2} + \frac{1}{2 s_0^2} \mathbb{E}_{\boldsymbol{\lambda} \cdot \boldsymbol{\gamma}}\left[\left(\lambda_j - \mathbf{z}_j^\intercal \boldsymbol{\gamma}\right)^2\right], \\
q^\star(\boldsymbol{\gamma}) &= \mathcal{N}_M\left(\boldsymbol{\mu}_\gamma^\star, s_0^2 \boldsymbol{\Sigma}_\gamma^\star\right), \\
&\quad \boldsymbol{\mu}_\gamma^\star = \boldsymbol{\Sigma}_\gamma^\star \mathbf{Z}^\intercal \mathbb{E}_{\varphi^2 \cdot \boldsymbol{\gamma}}\left[\boldsymbol{\Phi}^{-2} \boldsymbol{\lambda}\right], \quad \boldsymbol{\Sigma}_\gamma^\star = \left(\mathbf{Z}^\intercal \mathbb{E}_{\varphi^2}\left[\boldsymbol{\Phi}^{-2}\right] \mathbf{Z} + s_0^2 \mathbb{E}_{\kappa^2}\left[\mathbf{D}_\kappa^{-1}\right]\right)^{-1}, \\
q^\star\left(\kappa_d^2\right) &= \mathcal{IG}\left(e_d^\star, f_d^\star\right), \quad d = 1, \ldots, D, \\
&\quad e_d^\star = a_d + \frac{m_d}{2}, \quad f_d^\star = b_d + \frac{1}{2} \mathbb{E}_\gamma\left[\boldsymbol{\gamma}_d^\intercal \boldsymbol{\gamma}_d\right], \\
q^\star\left(\tau^2\right) &= \mathcal{IG}\left(\frac{p}{2} + 1, g^\star\right), \\
&\quad g^\star = \mathbb{E}_\zeta\left[\zeta^{-1}\right] + \frac{1}{2} \mathbb{E}_{\beta_0 \cdot \lambda_0^2 \cdot \tau^2}\left[\frac{\beta_0^2}{\tau^2 \lambda_0^2}\right] + \frac{1}{2} \sum_{j=1}^p \mathbb{E}_{\beta \cdot \sigma^2 \cdot \lambda}\left[\frac{\beta_j^2}{\sigma^2 \lambda_j^2}\right],
\end{aligned}$$

$$
\begin{aligned}
q^{\star}(\zeta) &= \mathcal{IG}\left(1, h^{\star}\right), \\
h^{\star} &= 1 + \mathbb{E}_{\tau^2}\left[\tau^{-2}\right], \\
q^{\star}\left(\sigma^2\right) &= \mathcal{IG}\left(v + \frac{n+p+1}{2}, l^{\star}\right), \\
l^{\star} &= q + \frac{1}{2}\left(\mathbb{E}_{\beta}\left[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2\right] + \mathbb{E}_{\beta_0 \cdot \lambda_0^2 \cdot \tau^2}\left[\frac{\beta_0^2}{\tau^2 \lambda_0^2}\right] + \sum_{j=1}^{p} \mathbb{E}_{\beta \cdot \lambda \cdot \tau^2}\left[\frac{\beta_j^2}{\tau^2 \lambda_j^2}\right]\right),
\end{aligned}
$$

其中，期望是相对于更新后的变分族 $q^{\star}(\boldsymbol{\theta})$ 计算的。边缘分布 $p(\mathbf{y})$ 的变分下界 $\mathcal{L}$ 计算如下:

$$
\begin{aligned}
\mathcal{L} \propto & \frac{1}{2}\left(\log\left|\boldsymbol{\Sigma}_{\beta}^{\star}\right| + \log\left|\boldsymbol{\Sigma}_{\gamma}^{\star}\right|\right) + \frac{p+1}{2}\mathbb{E}_{\sigma^2}\left[\log \sigma^2\right] - \left(v + \frac{n+p+1}{2}\right)\log l^{\star} + \\
& \sum_{j=1}^{p}\left(\log s_j - \log k_j + a_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j^{-2}\right] + b_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j^2\right] - c_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j\right] - \log d_j^{\star}\right) - \\
& \sum_{d=1}^{D} e_d^{\star}\log f_d^{\star} - \left(\frac{p}{2} + 1\right)\log g^{\star} - \log h^{\star} - \log a_0^{\star} - \log k_0^{\star}, \qquad (5)
\end{aligned}
$$

其中 $k_j = 1 - \mathbb{P}\left(\mathcal{N}\left(\mathbf{z}_j^{\intercal}\boldsymbol{\gamma},\ s_0^2\varphi_j^2\right) < 0\right)$ 且 s_j 是式（4）中参数 λ_j 的未知归一化常数。各组成部分的推导见补充材料的第 D.1 节。

与吉布斯采样器不同，变分算法无需显式求逆 $p \times p$ 矩阵 $\boldsymbol{\Sigma}_{\beta}^{\star}$，因为只需要计算 $\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{X}^{\intercal}\mathbf{y}$、对角矩阵 $\left(\boldsymbol{\Sigma}_{\beta}^{\star}\right)$ 和 $\mathbf{X}\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{X}^{\intercal}$。根据 M 的研究，这些量可以高效计算，复杂度为 unch 等人（2019年）。代数推导细节见补充材料的第E.1 节。算法的瓶颈在于计算 $\log s_j$、$\mathbb{E}_{\lambda}\left[\lambda_j\right]$、$\mathbb{E}_{\lambda}\left[\lambda_j^2\right]$ 和 $\mathbb{E}_{\lambda}\left[\lambda_j^{-2}\right]$，因为这些量没有闭式形式。可按照补充材料第 C.1 节所述程序并行计算这些量。

5 模拟研究

在本节中，我们通过实证展示变分近似联合后验分布的质量，以及利用协变量数据进行变量选择的优势。我们采用了一个类似于Johnson (2013提出的模拟方案。具体而言，我们在低维至中等维度的 p 问题上评估近似方法的有效性，而在更高维的情形下评估变量选择效果。有关数据生成方案的更多细节，请参阅《补充材料》第F.1节。

通过所选变量的ROC曲线下面积（AUC）的平均值来评估变量选择，其中一般协变量j的后验纳入概率通过量 $\mathbb{E}_{\lambda_j}\left[\lambda_j^2/\left(1+\lambda_j^2\right)\right]$ (Carvalho 等人，2010年）进行评估。

σ^2 和 κ_1^2 的超参数分别设为（v，q）=（1，10）和（a_1，b_1）=（1，10）。吉布斯采样器运行 $B=5000$ 次迭代，预热期为$bn=2500$ 次迭代；变分算法则在下界增幅小于 $\epsilon=0.001$ 或达到最大迭代次数$B=1000$ 时停止。

1. 变量选择性能。
在此，我们分析模型表现如何随先验信息程度的变化而变化。我们研究 $n=$（50，100，150）和 $p=$（500，1000，1500）这些情况，并将真实非零系数的数量设为 $p_0=30$。为研究模型如何从协变量数据中借用信息，我们模拟了代表五种先验信息程度的协变量数据源：

$G0$) **无协变量数据**设置：协变量数据矩阵中仅包含截距，因此$\mathbf{Z}=\mathbf{1}_p$;

$G1$) **非信息性**设置：通过随机选择 100 个回归变量，将一个二元协变量数据源纳入模型，因此协变量数据矩阵$\mathbf{Z}$ 由二元向量$\mathbf{z}\in\{0,1\}^p$构建，其中 $z_j=1$ 若选择第 j 个变量，则取值为 $z_j=0$，否则；

$G2$) **弱信息性**设置：通过随机选择 20 个真实非零回归变量和 80 个真实零回归变量，将一个二元协变量数据源纳入模型，因此协变量数据矩阵$\mathbf{Z}$ 由二元向量$\mathbf{z}\in\{0,1\}^p$构建，其中 $z_j=1$ 若选择第 j 个变量，则取值为 $z_j=0$，否则；

$G3$) **信息性**设置：通过随机选择 20 个真实非零回归变量和 10 个真实零回归变量，将一个二元协变量数据源纳入模型，因此协变量数据矩阵$\mathbf{Z}$ 由二元向量$\mathbf{z}\in\{0,1\}^p$构建，其中 $z_j=1$ 若选择第 j 个变量，则取值为 $z_j=0$，否则；

$G4$) **完美协变量数据信息** 设置：协变量数据矩阵 $\mathbf{Z}$由二元向量 $\mathbf{z}\in\{0,1\}^p$创建，其中 $z_j=1$ 若满足 $\beta_j^0\neq 0$，$z_j=0$ 否则。

结果如图1所示。模型能够从辅助信息中学习；当协变量数据确实包含信息时，变量选择性能会有所提升。随着先验信息强度的增加，AUC也随之提高；当协变量数据提供了完美信息时，变量便能被完美选出（病例 $G4$）。另一方面，随机协变量数据会导致性能略有下降，并使 AUC得分降低。具体而言，若比较随机协变量数据的情形（$G1$）与无协变量数据的情形（$G0$），后者的表现略好一些。不过，这种

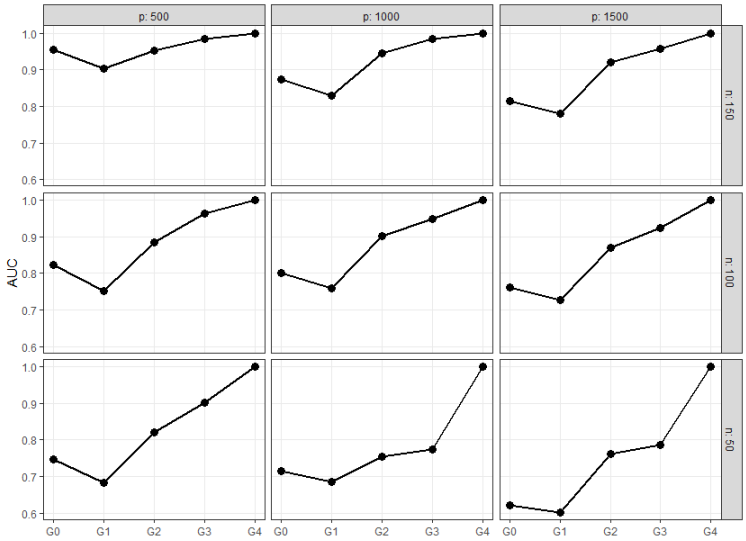


图1：在协变量数据情景 $G0$ 到 $G4$下使用变分算法进行变量选择；AUC 为实验中 25 次重复实验的均值。

差异会随着协变量数量的增加而消失。最后，正如预期，在 $G2$ 和 $G3$之间， AUC的增幅最小，因为这两种情形所代表的协变量信息程度最为相似。

2. 吉布斯采样器与变分推断。
我们通过比较变分近似与吉布斯采样器在变量选择和均方误差（MSE）方面的表现，评估变分近似的准确性$\boldsymbol{\beta}$。我们分析了病例 $n = (50, 100)$ 和 $p = (75, 125, 200)$，并设定$p_0 = 30$。我们考虑了四种先验信息程度，关于协变量数据模拟的更多细节，请参见补充材料第 F.2 节。

变量选择结果见补充材料中的图5。在所有考察的病例中，两种方法表现相似。具体而言，在维度最低的病例中，结果完全相同；而随着 p 增加，变分近似的表现逐渐接近吉布斯采样器。表1报告了两种方法在 $\boldsymbol{\beta}$ 估计方面的相似程度。吉布斯抽样和变分算法对回归参数给出了相似的估计，因为在所有考察的病例中， MSE 的均值都接近于零。特别是，当 n 较大且协变量数据提供的信息很强时，两种方法给出的估计几乎相同。

在补充材料的图6中，我们展示了不同协变量信息病例下回归参数标准差估计值的平均值 $\boldsymbol{\beta}$ 。结果与MacKay 等人(2003年); Wang 和 Titterington (2005年); Giordano 等人(2017年)的观点一致，因为与吉布斯抽样相比，变分近似会低估后验方差。然而，正如

表 1：在吉布斯抽样与变分算法之间的 $MSE_{\boldsymbol{\beta}}$ 均值；该均值基于每项实验的 10 次重复实验计算。

$MSE_{\boldsymbol{\beta}} = \|\boldsymbol{\beta}_{\mathrm{FB}} - \boldsymbol{\beta}_{\mathrm{GS}}\|_2^2$				
	$p = 75$	$p = 125$	$p = 200$	
$G0$	0.0079	0.2592	0.0365	$n = 50$
$G1$	0.0074	0.1034	0.0070	
$G2$	0.0091	0.0647	0.0621	
$G3$	0.0032	0.1153	0.0787	
$G0$	0.0006	0.0268	0.0217	$n = 100$
$G1$	0.0007	0.0023	0.0212	
$G2$	0.0005	0.0046	0.0023	
$G3$	< 0.0001	0.0037	0.0001	

Blei 和 Jordan (2006年)以及模拟结果所示，这种精度不足未必会影响模型的整体性能。

吉布斯采样器有一个主要缺点：在某些 n 和 p设定下，第B.1 节补充材料中介绍的抽样方法接受概率较低，因此吉布斯采样器的效率会受到严重影响。鉴于变分近似取得了良好结果，而吉布斯采样器在计算效率方面表现不佳，对于以下大型 p 应用，我们采用前者。

6 真实数据应用

本节介绍两个基因组学应用。第一个应用用于评估我们的方法在多元线性回归情境下处理二元协变量数据的效果。第二个应用涉及分类情境下的多个协变量数据源，且变量数量众多。

在第5 节中，我们表明，当协变量独立抽样时，阈值化变量选择效果良好。然而，后验概率是分别处理的，且最优阈值几乎从不为。5。因此，本节还将讨论一种较新的选择程序，称为解耦收缩与选择 （DSS），由Hahn 和 Carvalho （2015）提出。该方法旨在应对可能非常强的相关性，这种情况在许多基因组学数据集中都存在。作者提出了一种基于预测变量后验均值的后验变量选择摘要，由此得到一系列稀疏模型。任何先验分布都可用于实现收缩，并且收缩与基于后验分布的选择方法相互“解耦”。DSS 方法依靠优化损失函数，在预测误差与解的稀疏性ˆ之间进行权衡。给定估计值 $\boldsymbol{\beta}$，可通过自适应LASSO得到如下解：

求解以下优化问题，

$$\boldsymbol{\theta}^{DSS} = \text{argmin}_{\boldsymbol{\theta}} \frac{1}{n}\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\theta}\|_2^2 + \lambda \sum_{j=0}^{p} \frac{|\theta_j|}{|\hat{\beta}_j|},$$

其中，平滑参数 λ 充当阈值参数，可通过在一组取值上进行交叉验证（网格搜索）来估计。作者认为，与阈值化相比，该方法更具优势，主要是因为它能自然地处理多重共线性。

6.1 案例研究 1：p38MAPK 通路

根据Kpogbezan 等人 (2019年)，我们在 p38MAPK 通路数据集上测试该模型。我们研究单核苷酸多态性（SNPs）对该通路中基因的影响。我们使用了 GEUVADIS 项目（Lap- palainen 等人，2013）收集的一部分数据。对于 99 个基因中的每一个，我们收集了不同数量 p_t 的单核苷酸多态性，$t = 1, \ldots\ldots, 99$，最少为 $p_t = 56$，最多为 $p_t = 1169$。经过初步筛选后，从千人基因组计划中获得了 $n = 373$ 个 RNA 测序样本。由于人们认为位于基因范围内的单核苷酸多态性对基因表达的影响更大，因此分析中纳入了二元协变量数据源 $\mathbf{z}_t \in \{0,1\}^{p_t}$，其中 $z_{t,\ j} = 1$ 表示与基因 t 相关的第 j 个单核苷酸多态性位于该基因范围内，而 $z_{t,\ j} = 0$ 则表示不在其中。

σ^2 和 κ_1^2 的超参数设为 $(v,\ q) = (1,\ 10)$ 和 $(a_1,\ b_1) = (1,\ 10)$。当变分下界的增幅小于 $\epsilon = 0.001$，或达到最大迭代次数 $B = 1000$ 时，算法停止。

为评估模型的预测性能，将数据划分为训练集（$n_1 = 249$）和测试集（$n_2 = 124$）。我们使用全部 99 个基因的均方误差相对降幅（rrMSE）进行评估。第 t 个回归模型的 rrMSE 定义为

$$\text{rrMSE}_t = 1 - \frac{\text{MSE}_t}{\text{MSE}_0},$$

其中，MSE_0 和 MSE_t 分别为零模型的均方误差和第 t 个线性模型的均方误差。$\mathbf{y}$ 的均值以零为中心，因此我们在零模型中不包含截距。请注意，rrMSE_t 值越大，预测性能越好。相反，负值表示其预测性能不如零模型。

我们将所提出的方法与 LASSO 回归（Tibshirani，1996年）进行比较。不同稀疏程度下的结果如图 2 所示。对于 infHS 回归，采用第5节中的阈值化程序（浅灰色圆点）和 DSS 方法（黑色圆点）筛选单核苷酸多态性。与基于阈值化的变量选择相比，后者对所有基因的预测效果更好。这是由单核苷酸多态性之间的强相关性所致。另一方面，在比较 infHS 回归与 LASSO 回归时，Kpogbezan 等人 （2019年）提出的看法仍然成立。LASSO 的结果噪声更大，这可能是因为其对接近零的系数收缩较弱，

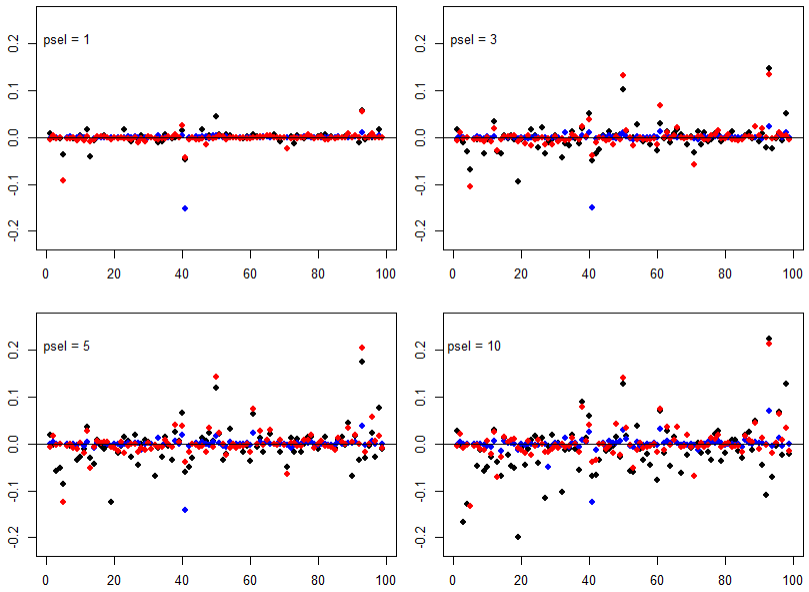


图 2：全部 99 个基因的 LASSO（深灰色圆点）、采用 DSS 选择程序的 infHS 回归（黑色圆点）和采用阈值选择程序的 infHS 回归（浅灰色圆点）的均方误差相对降幅（rrMSE）。每个面板中，选定的 SNP 数量上限分别固定为 1、3、5 和 10。

而 infHS 模型的大多数均方误差相对降幅$_t$ 都集中在零附近。单核苷酸多态性的影响只与少数基因相关：均方误差相对降幅$_t$ 最大的对应基因为 50、61 和 93。infHS 和 LASSO 都能捕捉到这些影响，但后者更容易出现负的均方误差相对降幅$_t$。对于基因 98，LASSO 的结果更好。请注意，Kpogbezan 等人 （2019年）的方法与我们的结果相当相似，因为该方法同样基于马蹄先验。不过，该方法的适用范围要窄得多，因为它只能处理一个离散型协变量数据源，且仅适用于连续型结局。

我们的方法用 68 分钟估计全部 99 个回归模型，并评估不同的变量选择方法。该算法在 x64 Windows 11 操作系统上运行。

6.2 案例研究 2：甲基化数据

该模型在Verlaat 等人（2018的甲基化数据集上进行测试。该数据集包含自采集宫颈阴道灌洗液的甲基化谱，样本来自28名宫颈正常女性和36名患有高级别癌前病变（宫颈上皮内瘤变3级）的女性，总计 $n = 64$ 个样本。为提高诊断分类效果，我们在分析中纳入 $D = 5$ 个协变量数据源：探针的标准差、先前研究中的 p 值、用于区分低甲基化和

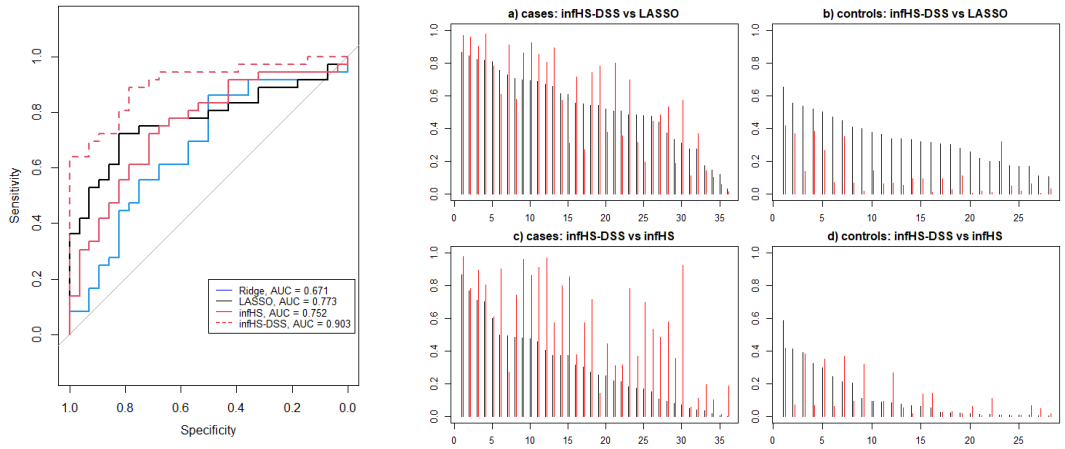


图3：留一交叉验证的结果。左图：岭回归（浅灰色）、LASSO（深灰色）、信息性马蹄回归（黑色）以及采用DSS变量选择程序的信息性马蹄回归（黑色虚线）的ROC曲线；右图：a 和b. 采用 infHS-DSS（黑色）和 LASSO（深灰色）得到的病例（$y_i = 1$）与对照组（$y_i = 0$）的预测概率，按 LASSO 预测值降序排列；c 和d. 采用 infHS-DSS （黑色）和普通 infHS （深灰色）得到的病例与对照组的预测概率，按 infHS 预测值降序排列。

高甲基化探针的二元变量、一个包含6个类别且表示每个探针所属基因组区域的分类变量，以及癌细胞中的探针均值。关于 Probit回归模型中联合后验分布的变分近似的理论细节，请参见补充材料的 C.2 节。根据先前的结果，我们考虑外部 p 值 $p_\alpha < 0.005$的探针，共计 $p = 11251$ 个探针；每个探针测量基因组上一个独特位置的甲基化水平。

我们将模型与普通岭回归和 LASSO 进行比较。对于 $d = 1, \ldots, 5$，超参数设为 $(a_d, b_d) = (1, 10)$，并通过留一交叉验证（LOOCV）评估预测性能。与病例研究1相同，若变分下界的增量小于 $\epsilon = 0.001$，或达到最大迭代次数 $B = 1000$，则停止算法。分析中还纳入了 infHS 模型的阈值化版本和 DSS 版本。前者不会从模型中排除任何协变量，因为所有后验纳入概率均大于 0.5。因此不展示这种方法的结果。我们还拟合了不使用协变量数据的模型；由于有许多病例（$y_i = 1$）对应的预测概率极低，其结果不及 infHS，因此此处不予展示。

图3 左侧面板展示了所考察竞争方法的ROC曲线，其中 infHS 的预测概率计算方式为 $p_i = \Phi(\mathbf{x}_i \boldsymbol{\beta})$。infHS-DSS 的预测结果最佳，曲线下面积为 $= 0.903$，较

表2：真实标签（$y_i = 0$ 或 $y_i = 1$）与预测概率（$p_i = \Phi(\mathbf{x}_i^\intercal \boldsymbol{\beta})$）之间的平均绝对差。

	岭回归	LASSO	infHS	infHS-DSS
$\sum_{i=1}^{64} \frac{\lvert y_i - p_i \rvert}{64}$	0.466	0.410	0.444	**0.299**

其普通版本的曲线下面积 $= 0.752$有显著提升；LASSO 的表现优于 infHS，曲线下面积为 $= 0.773$，并给出了最稀疏模型，平均纳入 8 个变量，而 infHS-DSS 估计纳入 49 个。在这 49 个探针中，有 37 个在超过 70% 的折中被选中，而只有 10 个在所有折中都被选中。普通岭回归在预测性能方面表现不佳（曲线下面积 $= 0.671$）。值得注意的是，GRridge 协变量数据方法（Van de Wiel 等人，2016）是少数同样能够处理多个协变量数据源的方法之一；此前据报道，其曲线下面积为 $= 0.77$ (Verlaat 等人，2018）。

图3右侧面板展示了 infHS-DSS 模型与 LASSO 和普通 infHS 的预测概率。与 LASSO 相比（图a和b，数值按 LASSO 预测值递减排列），infHS-DSS 对病例给出的预测结果相近（$y_i = 1$），其中一些观测的预测得分较高，另一些则较低；另一方面，它为对照组给出的得分显著更低（$y_i = 0$），从而降低了假阳性率并提升了整体表现。我们可以通过分析图c 和d（数值按 infHS 预测值递减排列），评估 DSS 程序对预测的影响。这两幅图对比了 infHS-DSS 与 infHS 模型。前者的表现远胜后者，因为它对几乎所有病例都给出了更高的概率，并降低了对得分最高的对照组给出的预测概率。为总结这些结果，我们计算了真实标签与预测概率之间的平均绝对差。结果如表2所示，其中 infHS-DSS 的表现最佳。

我们在表 3 中总结了协变量信息，其中展示了总体协变量数据源的分布以及 $p_{sel} = 37$ 最常选中的探针的分布。与癌细胞中探针均值相关的协变量数据源，在原始特征与选中特征之间的分布差异最大，这表明该数据源在提供先验信息方面最具信息量。其他协变量数据源未发现明显差异。

infHS 的估计计算时间约为 5 分钟。该方法在 x64 Windows 11 操作系统上进行测试，局部方差通过 4 个核心并行评估。

7 讨论

我们提出了一种新颖的贝叶斯回归方法，适用于能够利用辅助先验信息（即协变量数据）进行学习的高维数据。我们证明，预测和

表3：在 $p_{sel} = 37$个最常选中的探针和原始总体中，协变量数据估计值及协变量数据分布的汇总。

协变量数据	选定探针分布 ($p_{sel} = 37$)	原始分布 ($p = 11251$)
外部 p-值	均值：0.0016 标准差：0.0014 范围：($2.4 \cdot 10^{-6}$,0.005)	均值：0.0019 标准差：0.0015 范围：($9.9 \cdot 10^{-11}$，0.005)
探针 标准差	均值：0.035 标准差：0.022 范围：(0.014,0.129)	均值：0.038 标准差：0.019 范围：(0.009,0.320)
基因组区域	远端：35.1% 岛：43.2% N 岸：5.4% 北部近岸 10.8% 南部陆架：0% 南部近岸：5.4%	远端：31.7% 岛：37.3% N 岸：4.6% 北部近岸：12.4% 南部陆架：4% 南部近岸：10%
甲基化程度	低甲基化：57% 高甲基化：43%	低甲基化：56% 高甲基化：44%
癌细胞均值	均值：0.150 标准差：1.116 范围：(-1.28,3.34)	均值：-0.099 标准差：1.137 范围：(-4.89,8.10)

当协变量数据确实具有信息量时，将其纳入模型有助于预测结局和进行变量选择。该模型既适用于连续型和二元结局，也适用于连续型和离散型协变量数据源。具体而言，我们提出了一种灵活的方法，可联合估计多个协变量数据系数；这与Van Nee 等人(2021年提出的、分别对每个数据源建模的方法不同。

我们讨论了一种完整的贝叶斯方法，并开发了吉布斯采样器，通过从各参数的全条件分布中抽样，迭代更新每个参数。为此，我们提出了一种新颖的拒绝采样方法，用于对局部方差进行抽样，因为其目标密度没有闭合形式。最后，我们提出了联合后验分布的变分近似，并应用坐标上升变分推断算法优化目标密度。后一种方法适用于高维问题，因为它不需要显式求逆 $p \times p$后验协方差矩阵 $\mathbf{\Sigma}^{\star}_{\beta}$。具体而言，只需其对角元素；我们采用了M¨unch 等人(2019年提出的方法来高效地实现这一点。此外，另一项计算优势是可以并行计算局部方差。对这些参数的计算是该方法的计算瓶颈：在某些参数设置下，拒绝采样的接受概率较低；而变分近似所需的数值积分则是算法中最耗费计算资源的步骤。

在后验推断方面，变分推断不如吉布斯采样器有用，因为它提供的是后验均值点估计，并且会低估后验方差（MacKay 等人，2003；Wang 和 Titterington，2005年；Turner 和 Sahani，2011年；Giordano 等人，2017年）。另一方面，这种精度不足不一定会影响模型的预测性能（Blei 和 Jordan，2006年）。因此，对于大型 p 数据集，应使用变分算法；而当关注后验推断（可信区间）且协变量数量适中时，吉布斯抽样可能会很有用。

对于某些应用而言，infHS 的一个潜在局限是：当分类协变量数据源中包含一个规模相对较小、但效应非常强的协变量数据组时，在这种情况下，对该特定变量组采用稀疏性假设可能并不现实。对于这类应用，infHS 的一个有趣扩展是允许对一小组变量采用稠密先验，因为预期这些变量具有特别相关的先验信息。

另一种提高灵活性的扩展方式是指定先验$\lambda_j \sim \text{Half-}t(v)$，其中 $v > 1$。这里 $v = 1$ 是因为目的是应用马蹄先验。在 Biswas 等人 (2021年)中，作者针对这种特定先验为高维回归提出了耦合技术，并指出，较大的 v 值会影响模型的统计性能和计算性能。

总而言之，我们提供了一种速度最快的实现方法之一，可利用辅助信息进行马蹄回归。R 代码`infHS` 以及所用数据集均可在以下网址获取：https://github.com/cbusatto/infHS。

致谢

第一作者还得到了 THE—托斯卡纳健康生态系统项目的支持。作者感谢 Francesco C. Stingo 教授在项目启动阶段提供的帮助。

参考文献

Albert, J. H. 和 Chib, S.（1993）。二元和多分类响应数据的贝叶斯分析。美国统计学会会刊，88(422):669–679。Beal, M. J.（2003）。用于近似贝叶斯推断的变分算法。伦敦大学，伦敦大学学院（英国）。Bhattacharya, A.、Chakraborty, A. 和 Mallick, B. K.（2016）。采用高斯尺度混合先验进行高维回归的快速采样。*Biometrika*，103 4:985–991。Bishop, C. M. 和 Nasrabadi, N. M.（2006）。模式识别与机器学习，第4卷。施普林格。Biswas, N.、Bhattacharya, A.、Jacob, P. E. 和 Johndrow, J. E.（2021）。基于耦合的收敛性评估：针对采用收缩先验的高维贝叶斯回归的若干吉布斯采样器。英国皇家统计学会会刊。*B*辑（统计方法学）。Blei, D. M. 和 Jordan, M. I.（2006）。狄利克雷过程混合模型的变分推断。《贝叶斯分析》，1(1):121–143。Blei, D. M.、Kucukelbir, A. 和 McAuliffe, J. D.（2017）。变分推断：面向统计学家的综述。美国统计学会会刊，112(518):859–877。Boonstra, P. S.、Taylor, J. M. G. 和 Mukherjee, B.（2013）。结合辅助信息以改进高维数据集上的预测：收缩方法集成。《生物统计学》（牛津，英格兰），14(2):259–272。Busatto, C. 和 Stingo, F. C.（2023）。利用图形马蹄先验推断多个高维网络。Carvalho, C. M.、Polson, N. G. 和 Scott, J. G.（2010）。稀疏信号的马蹄估计量。*Biometrika*，97:465–480。Geweke, J.（1993）。独立 Student-t 线性模型的贝叶斯处理。应用计量经济学杂志，8:S19–S40。Giordano, R.、Broderick, T. 和 Jordan, M.（2017）。协方差、稳健性与变分贝叶斯。机器学习研究期刊，19。

Hahn, P. R. 和 Carvalho, C. M.（2015）。贝叶斯线性模型中收缩与选择的解耦：后验摘要视角。美国统计学会会刊，110(509):435–448。Johndrow, J. E.、Orenstein, P. 和 Bhattacharya, A.（2020）。针对马蹄先验的可扩展近似马尔可夫链蒙特卡罗算法。机器学习研究期刊，21:73:1–73:61。Johnson, V. E.（2013）。高维和超高维情形下贝叶斯模型选择的数值问题。《贝叶斯分析》（在线版），8(4):741。Jordan, M.、Ghahramani, Z.、Jaakkola, T. 和 Saul, L.（1999）。图模型变分方法导论。《机器学习》，37:183–233。Kpogbezan, G.、van de Wiel, M.、van Wieringen, W. 和 van der Vaart, A.（2019）。在高维稀疏回归中利用马蹄先验和变分贝叶斯纳入先验信息并借用信息。Lappalainen, T.、Sammeth, M.、Friedl¨ander, M.、Hoen, P.、Monlong, J.、Rivas, M.、Gonz`alez Porta, M.、Kurbatova, N.、Griebel, T.、Ferreira, P.、Barann, M.、Wieland, T.、Greger, L.、Iterson, M.、Alml¨of, J.、Ribeca, P.、Pulyakhina, I.、Esser, D.、Giger, T. 和 Dermitzakis, E.（2013）。转录组和基因组测序揭示人类的功能变异。《自然》，501。Lee, S. Y.（2022）。吉布斯采样器与坐标上升变分推断：集合论综述。《统计学通讯：理论与方法》，51(6):1549–1568。MacKay, D. J.、Mac Kay, D. J. 等（2003）。信息论、推断与学习算法。剑桥大学出版社。Makalic, E. 和 Schmidt, D. F.（2016）。一种简单的马蹄估计量采样器。《*IEEE*信号处理快报》，23(1):179–182。M¨unch, M.、Peeters, C.、van der Vaart, A. 和 van de Wiel, M.（2019）。自适应组正则化逻辑弹性网络回归。《生物统计学》（牛津，英格兰），22。Neuenschwander, B.、Roychoudhury, S. 和 Schmidli, H.（2016）。临床试验中协变量数据的使用。统计与生物制药研究，8(3):345–354。Polson, N. G.、Scott, J. G. 和 Windle, J.（2013）。利用 P´olya–Gamma 潜变量进行逻辑模型的贝叶斯推断。美国统计学会会刊，108(504):1339–1349。Salimans, T.、Kingma, D. P. 和 Welling, M.（2015）。马尔可夫链蒙特卡罗与变分推断：弥合差距。载于第*32*届国际机器学习会议论文集——第*37*卷，ICML’15，第1218–1226页。

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R.（2013）。稀疏组 Lasso。计算图形统计学杂志, 22(2):231–245。Tai, F. and Pan, W.（2007）。在包含多个惩罚项的惩罚分类器中纳入预测变量的先验知识。生物信息学, 23(14):1775–1782。Tibshirani, R.（1996）。通过 Lasso 实现回归收缩与变量选择。皇家统计学会杂志 *B*辑（方法论）, 58(1):267–288。Turner, R. E. and Sahani, M.（2011）。时间序列模型变分期望最大化的两个问题。载于贝叶斯时间序列模型，第109–130页。剑桥大学出版社。Van de Wiel, M. A., Lien, T. G., Verlaat, W., van Wieringen, W. N., and Wilting, S. M.（2016）。利用协变量数据改进预测：自适应组正则化岭回归。医学统计学, 35(3):368–381。Van de Wiel, M. A., Te Beest, D. E., and M¨unch, M. M.（2019）。从大量数据中学习：用于高维模型预测的经验贝叶斯。斯堪的纳维亚统计学杂志, 46(1):2–25。Van Nee, M. M., Wessels, L. F., and van de Wiel, M. A.（2021）。用于高维预测的灵活协变量数据学习。医学统计学, 40(26):5910–5925。Velten, B. and Huber, W.（2019）。利用变分贝叶斯和外部协变量进行高维回归与分类中的自适应惩罚。《生物统计学》, 22(2):348–364。Verlaat, W., Snoek, B. C., Heideman, D. A., Wilting, S. M., Snijders, P. J., Novianti, P. W., van Splunter, A. P., Peeters, C. F., van Trommel, N. E., Massuger, L. F., et al.（2018）。鉴定并验证用于基于 HPV 的宫颈筛查的三基因甲基化分类器，样本来自自采样本。临床癌症研究, 24(14):3456–3464。Wang, B. and Titterington, D. M.（2005）。变分贝叶斯近似所对应的区间估计不足。载于第十届人工智能与统计国际研讨会，《机器学习研究论文集》，第373–380页。机器学习研究论文集。Yuan, M. and Lin, Y.（2006）。含分组变量回归中的模型选择与估计。皇家统计学会杂志 *B*辑, 68:49–67。

A 网络附录 A

A.1 局部方差的吉布斯步骤 λ_0^2

线性回归模型中，截距 β_0 的局部收缩参数 λ_0 所采用的半柯西先验与正态误差下的方差不共轭。我们依赖

Makalic 和 Schmidt 提出的数据增广步骤Makalic 和 Schmidt (2016)，以便轻松高效地更新参数 λ_0。

命题 4。若 $\lambda_0^2 \mid \psi_0 \sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\psi_0}\right)$ 且 $\psi_0 \sim \mathcal{IG}\left(\frac{1}{2}, 1\right)$，则 $\lambda_0 \sim \mathcal{C}^+ \quad (0,1)$ (Makalic 和 Schmidt，2016)。

逆伽马分布与其自身以及局部尺度参数共轭，因此可以得到闭式形式的全条件分布，并执行吉布斯步骤。λ_0^2 和 ψ_0 的全条件分布为

$$\lambda_0^2 \mid \beta_0, \tau^2, \sigma^2 \sim \mathcal{IG}\left(1, \frac{1}{\psi_0} + \frac{\beta_0^2}{2\sigma^2\tau^2}\right),$$
$$\psi_0 \mid \lambda_0^2 \sim \mathcal{IG}\left(1, 1 + \frac{1}{\lambda_0^2}\right).$$

B 网络附录 B

B.1 参数 λ_j的拒绝采样

本节提出一种新颖的拒绝采样算法，在未知归一化常数的情况下，从全条件分布中对收缩参数 λ_j、$j = 1$、……、p进行采样。拒绝采样通过从提议分布 g 中抽样，并以与比值 $r(x_\star) = f(x_\star)/g(x_\star)$成正比的概率接受该样本，从而从密度 f 中抽取新值 $x_\star$。具体而言，必须选择 g，使得 f 的支撑集是 g支撑集的子集。令

$$k = \sup_x \frac{f(x)}{g(x)} < \infty$$

并以概率 $a = f(x_\star)/(kg(x_\star))$ 接受 $x_\star$。可以证明，该算法的接受概率为 $1/k$。因此，目标是找到一个提议分布 g，使得 k 较小。只要有对 g 进行采样的方法，即使 f 的归一化常数未知，拒绝采样仍然适用。

考虑以下密度

$$f(x) = c_f x^{-1} e^{-\psi/x^2 - \alpha^2 x^2 + \beta x} \cdot \mathbb{I}_{(x>0)},$$

其中 ψ，$\alpha^2 > 0$，$\beta \in \mathbb{R}$ 且 c_f 是未知的归一化常数。根据提议分布 $g(x) \sim \mathcal{G}_{3p}(\gamma, \alpha, \beta)$ 抽取新值，其密度为 $g(x) = c_g x^\gamma e^{-\alpha^2 x^2 + \beta x}$，其中 $\gamma \in \mathbb{N}^+$。我们令参数 α 和 β 在 f 和 g 中相等，以便比率 $r(x)$ 在解析上易于处理，且容易最大化。最终，通过选择参数 γ来优化接受概率。选择 α 和 β 可得到

$$r(x) = \frac{c_f}{c_g} x^{-\gamma-1} e^{-\psi/x^2} \cdot \mathbb{I}_{(x>0)},$$

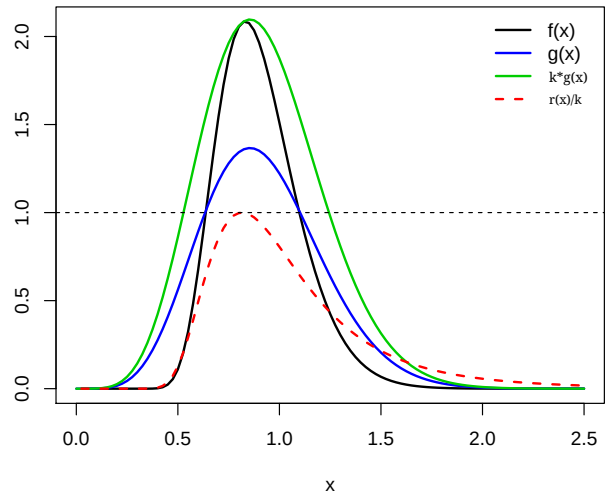


图4： $f(x)$与 $g(x)$在 $\psi = 2$、 $\alpha^2 = 2.25$ 和 $\beta = -2$下的比较。估计参数为 $\gamma = 5$；接受概率为 $a = 0.65$。

其在 $\dot{x} = \sqrt{2\psi/(\gamma + 1}$ 处有一个正最大值。关于从 $\mathcal{GIG}_{3p}$ 分布中抽样的理论性质和详细信息见 Busatto 和 Stingo (2023)。

选择参数 γ 的理论最优方法是对 $k = r(\dot{x})\, c_f/c_g$ 关于 γ 求导，并找出最优值。然而，这只能通过迭代方法实现，这会对该方法的计算效率产生负面影响。另一种解决方案是设置参数 γ，使分布 f 和 g 在同一数值 x_{max} 处达到最大值。也就是说，通过求解一个四次方程来计算 f 的最大值，并将参数 γ 估计为

$$\gamma = x_{max}\left(2\alpha^2 x_{max} - \beta\right) > -1.$$

由于 $\gamma \in \mathbb{N}_0^+$，因此选择最接近的非负整数。该方法的一个示例见补充材料中的图4，其中通过数值积分计算归一化常数 c_f。

对于参数 ψ、 α 和 β 的某些取值，所提算法的接受概率会逐渐降至 0，从而影响算法的效率。因此，为了避免显式求逆协方差矩阵 $\boldsymbol{\Sigma}_{\beta}^{\star}$，我们将在下一节介绍一种能够解决这些问题的变分贝叶斯算法。

C 网络附录 C

C.1 变分推断

变分推断（VI）是一种确定性优化方法，用于近似目标密度 $\pi(\boldsymbol{\theta} \mid \mathrm{y}, \mathrm{X}, \mathrm{Z})$，其中 $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda_0^2, \boldsymbol{\lambda}, \psi_0, \boldsymbol{\varphi}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \tau^2, \zeta, \sigma^2)$，并使用另一个变分分布 $q(\boldsymbol{\theta})$，从而将贝叶斯推断转化为一个优化问题（Salimans et al., 2015; Lee, 2022）。目标是找到使下式最小化的 $q(\boldsymbol{\theta})$

目标密度与变分分布之间的 Kullback-莱布勒散度（KL）。其中的最小化问题最终转化为最大化变分下界，其定义为 $\mathcal{L}=\mathbb{E}_q\left[\log\ \pi(\boldsymbol{\theta},\mathbf{y},\mathbf{X},\mathbf{Z})\right]-\mathbb{E}_q\left[\log\ q(\boldsymbol{\theta})\right]$。

$q(\boldsymbol{\theta})$的一种常见因子分解是所谓的平均场变分近似（Jordan 等人，1999；Beal，2003），这是在计算可处理性与性能准确性之间的一种折中。假设变分族 $q(\boldsymbol{\theta})$是相互独立的边缘变分因子$q_k(\theta_k)$、 $k=1$、……、K的乘积。我们采用坐标上升变分推断算法（CAVI）（Bishop 和 Nasrabadi，2006年；Blei 等人，2017年）来求解上述优化问题。在下界 $\mathcal{L}$收敛之前，CAVI 算法会基于先验分布的超参数以及因子 $q_{-k}(\theta_{-k})$当前的期望（视为固定）迭代更新变分因子 $q_k(\theta_k)$、 $k=1$、……、 K的参数（Lee，2022）。这样，模型便能够考虑参数之间的非线性依赖关系。在平均场近似下，各组成部分被假定为相互独立，此时最优解为

$$q^{\star}\left(\theta_k\right)\propto\exp\left\{\mathbb{E}_{q_{-k}}\left[\log\pi\left(\theta_k\mid\theta_{-k},\mathbf{y},\mathbf{X},\mathbf{Z}\right)\right]\right\}.$$

尽管因子之间相互独立的假设尤为严格，但坐标上升变分推断算法提供了一种灵活的方法，并且可以保证收敛到局部最优解（Blei 等人，2017年）。请注意，在共轭框架下使用指数族时，变分因子 $q(\theta_k$与可处理分布$\pi\left(\theta_k\mid\theta_{-k}\right.$，$\mathbf{y}$，$\mathbf{X}$，$\left.\mathbf{Z}\right)$具有相同的核。

变分下界的计算方式为

$$\begin{aligned}
\mathcal{L}&=\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\theta},\mathbf{y},\mathbf{X},\mathbf{Z}\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\theta}\right)\right]\\
&=\mathbb{E}_{q^{\star}}\left[\log p\left(\mathbf{y}\mid\mathbf{X},\boldsymbol{\beta},\sigma^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\beta}\mid\sigma^2,\tau^2,\lambda_0^2,\boldsymbol{\lambda}\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\lambda_0^2\mid\psi_0\right)\right]+\\
&\quad\mathbb{E}_{q^{\star}}\left[\log\pi\left(\psi_0\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\lambda}\mid\mathbf{Z},\boldsymbol{\gamma},\boldsymbol{\varphi}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\varphi}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\gamma}\mid\boldsymbol{\kappa}^2\right)\right]+\\
&\quad\mathbb{E}_{q^{\star}}\left[\log\pi\left(\boldsymbol{\kappa}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\tau^2\mid\zeta\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\zeta\right)\right]+\mathbb{E}_{q^{\star}}\left[\log\pi\left(\sigma^2\right)\right]-\\
&\quad\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\beta}\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\lambda_0^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\psi_0\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\lambda}\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\varphi}^2\right)\right]-\\
&\quad\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\gamma}\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\kappa}^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\tau^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\zeta\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\sigma^2\right)\right]\\
&\propto\frac{1}{2}\left(\log|\boldsymbol{\Sigma}_{\beta}^{\star}|+\log|\boldsymbol{\Sigma}_{\gamma}^{\star}|\right)+\frac{p+1}{2}\mathbb{E}_{\sigma^2}\left[\log\sigma^2\right]-\left(v+\frac{n+p+1}{2}\right)\log l^{\star}+\\
&\quad\sum_{j=1}^{p}\left(\log s_j-\log k_j+a_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j^{-2}\right]+b_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j^2\right]-c_j^{\star}\mathbb{E}_{\lambda}\left[\lambda_j\right]-\log d_j^{\star}\right)-\\
&\quad\sum_{d=1}^{D}e_d^{\star}\log f_d^{\star}-\left(\frac{p}{2}+1\right)\log g^{\star}-\log h^{\star}-\log a_0^{\star}-\log k_0^{\star},
\end{aligned}\tag{6}$$

其中 $k_j=1-\mathbb{P}\left(\mathcal{N}\left(\mathbf{z}_j\underset{\tau}{\boldsymbol{\gamma}},\ s_0^2\varphi_j^2\right)<0\right)$并且$s_j$是主文（4）中参数 λ_j 的未知归一化常数。各个分量的推导见补充材料第D.1 节。

与吉布斯采样器不同，变分算法不需要显式求逆 $p \times p$ 矩阵 $\boldsymbol{\Sigma}_{\beta}^{\star}$，因为只需要 $\boldsymbol{\Sigma}_{\beta}^{\star}$ $\mathbf{X}^{\intercal}\mathbf{y}$、对角矩阵 $\left(\boldsymbol{\Sigma}_{\beta}^{\star}\right)$和$\mathbf{X}^{\intercal}\mathbf{X}$ 这些量。按照 $\mathcal{O}\left(n^2 p\right)$的研究方法，这些量可以高效计算，计算复杂度为M¨unch 等（2019年）。详情见补充材料第E.1 节。该算法的瓶颈在于计算 $\log s_j$、 $\mathbb{E}_{\lambda}\left[\lambda_j\right]$、 $\mathbb{E}_{\lambda}\left[\lambda_j^2\right]$ 和 $\mathbb{E}_{\lambda}\left[\lambda_j^{-2}\right]$这些项，因为它们没有闭式形式。可以通过数值积分计算这些项，例如使用高斯求积公式或其自适应变体——高斯-克朗罗求积公式。根据后验参数 $a_j^{\star}$、 $b_j^{\star}$和 $c_j^{\star}$，这种策略容易出现数值不稳定。为避免溢出问题，我们按以下步骤计算这些积分：

Step 0: assume we have to evaluate $\int_0^{\infty} f(x)dx$, where $f(x) \propto x^{\nu} \exp\left\{-dx^{-2} - bx^2 + cx\right\} \cdot \mathbb{I}_{(\lambda_j > 0)}$, where $\nu = \{-3, -1, 0, 1\}$;

步骤 *1*： 通过求解一个四次方程计算最大值 $\dot{x}$ ，并在对数尺度上评估$f(\dot{x})$;

Step 2: evaluate $i_0 = \int_0^{\infty} \exp\left\{\log f(x) - \log f(\dot{x})\right\} dx$ numerically;

步骤 *3*： 返回 $\exp\left\{\log i_0 + \log f(\dot{x})\right\}$。

C.2 Probit 回归中的变分推断

令 $\mathbf{y} \in \{0,1\}^n$ 为一个二元向量，且 $\mathbf{X} \in \mathbb{R}^{n \times (p+1)}$ 为设计矩阵。根据Albert 和 Chib (1993年)，我们引入潜变量 $\mathbf{w} \in \mathbb{R}^n$ ，以构建用于更新回归系数向量的共轭框架 $\boldsymbol{\beta}$。Probit 模型的假设为

$$
\begin{aligned}
y_i \mid \mathbf{w} &= \begin{cases} 1 & \text{if } w_i > 0 \\ 0 & \text{if } w_i \leq 0, \end{cases} \\
w_i \mid \mathbf{X}, \boldsymbol{\beta} &\sim \mathcal{N}\left(\mathbf{x}_i^{\intercal}\boldsymbol{\beta}, 1\right), \quad i = 1, \ldots, n, \\
\beta_0 \mid \tau^2, \lambda_0^2 &\sim \mathcal{N}\left(0, \tau^2\lambda_0^2\right) \\
\beta_j \mid \tau, \lambda_j &\sim \mathcal{N}\left(0, \tau^2\lambda_j^2\right), \quad j = 1, \ldots, p.
\end{aligned}
$$

参数的先验分布 $\boldsymbol{\lambda}$、$\boldsymbol{\gamma}$ 和 τ 与第 2 节相同，算法也遵循主文第 3 节的相同步骤。引入潜变量 $\mathbf{w}$ 后，可将问题表述为关于潜变量的正态回归模型 w_i。回归参数向量 $\boldsymbol{\beta}$ 的正态先验分布与 $\mathbf{w}$的正态分布共轭。模型的联合后验分布为

$$
\pi\left(\boldsymbol{\theta}, \mathbf{w}, \mathbf{y}, \mathbf{X}, \mathbf{Z}\right) \propto p\left(\mathbf{y} \mid \mathbf{w}\right) \cdot \pi\left(\mathbf{w} \mid \mathbf{X}, \boldsymbol{\theta}\right) \cdot \pi\left(\boldsymbol{\theta} \mid \mathbf{Z}\right),
$$

其中 $\boldsymbol{\theta}=\left(\boldsymbol{\beta}、\ \lambda_0^2、\ \boldsymbol{\lambda}、\ \psi_0、\ \boldsymbol{\varphi}、\ \boldsymbol{\gamma}、\ \boldsymbol{\kappa}、\ \tau^2、\ \zeta\right)$。在平均场近似下，参数 $\mathbf{w}$ 和 $\boldsymbol{\beta}$ 的变分因子更新如下

$$
\begin{aligned}
q^{\star}\left(w_i\right) &= \begin{cases}\mathcal{N}_{+}\left(\mu_i^{\star}, 1\right) & \text{if } y_i=0 \\ \mathcal{N}_{-}\left(\mu_i^{\star}, 1\right) & \text{if } y_i=0,\end{cases} \\
q^{\star}(\boldsymbol{\beta}) &= \mathcal{N}_p\left(\boldsymbol{\mu}_\beta^{\star}, \boldsymbol{\Sigma}_\beta^{\star}\right), \\
\boldsymbol{\mu}_\beta^{\star} &= \boldsymbol{\Sigma}_\beta^{\star} \mathbf{X}^{\intercal} \mathbb{E}_w[\mathbf{w}], \quad \boldsymbol{\Sigma}_\beta^{\star}=\left(\mathbf{X}^{\intercal} \mathbf{X}+\mathbb{E}_{\lambda_0^2 \cdot \lambda \cdot \tau^2}\left[\tau^{-2} \boldsymbol{\Lambda}^{-2}\right]\right)^{-1},
\end{aligned}
$$

其中，$\mu_i^{\star}=\mathbf{x}_i \mathbb{E}_\beta[\boldsymbol{\beta}]$ 和 $\mathcal{N}_{+}$ 以及 $\mathcal{N}_{-}$ 分别表示在 0 处左截断和右截断的正态分布。潜变量 w_i 的期望为

$$
\mathbb{E}_w\left[w_i\right]=\begin{cases}\mu_i^{\star}+\frac{\phi_i}{1-\Phi_i} & \text{if } y_i=0 \\ \mu_i^{\star}-\frac{\phi_i}{\Phi_i} & \text{if } y_i=0,\end{cases}
$$

其中，$\phi_i=\phi\left(-\mu_i^{\star}\right)$ 和 $\Phi_i=\Phi\left(-\mu_i^{\star}\right)$ 分别为正态密度函数和正态累积分布函数。变分下界计算如下：

$$
\begin{aligned}
\mathcal{L} &= \mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\theta}, \mathbf{w}, \mathbf{y}, \mathbf{X}, \mathbf{Z}\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\boldsymbol{\theta}, \mathbf{w})\right] \\
&= \mathbb{E}_{q^{\star}}\left[\log p\left(\mathbf{y} \mid \mathbf{w}\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\mathbf{w} \mid \mathbf{X}, \boldsymbol{\beta}\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\beta} \mid \tau^2, \lambda_0^2, \boldsymbol{\lambda}\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\lambda_0^2 \mid \psi_0\right)\right]+ \\
&\quad \mathbb{E}_{q^{\star}}\left[\log \pi\left(\psi_0\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\lambda} \mid \mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\varphi}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\varphi}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\gamma} \mid \boldsymbol{\kappa}^2\right)\right]+ \\
&\quad \mathbb{E}_{q^{\star}}\left[\log \pi\left(\boldsymbol{\kappa}^2\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi\left(\tau^2 \mid \zeta\right)\right]+\mathbb{E}_{q^{\star}}\left[\log \pi(\zeta)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\mathbf{w})\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\boldsymbol{\beta})\right]- \\
&\quad \mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\lambda_0^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\psi_0\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\boldsymbol{\lambda})\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\varphi}^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\boldsymbol{\gamma})\right]- \\
&\quad \mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\boldsymbol{\kappa}^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}\left(\tau^2\right)\right]-\mathbb{E}_{q^{\star}}\left[\log q^{\star}(\zeta)\right] \\
&\propto \sum_{i=1}^n\left(y_i \log\left(1-\Phi_i\right)+\left(1-y_i\right) \log\left(\Phi_i\right)\right)+\frac{1}{2}\left(\log\left|\boldsymbol{\Sigma}_\beta^{\star}\right|+\log\left|\boldsymbol{\Sigma}_\gamma^{\star}\right|\right)+ \\
&\quad \sum_{j=1}^p\left(\log s_j-\log k_j\right)+\sum_{j=1}^p\left(a_j^{\star} \mathbb{E}_\lambda\left[\lambda_j^{-2}\right]+b_j^{\star} \mathbb{E}_\lambda\left[\lambda_j^2\right]-c_j^{\star} \mathbb{E}_\lambda\left[\lambda_j\right]-\log d_j^{\star}\right)- \\
&\quad \sum_{d=1}^D e_d^{\star} \log f_d^{\star}-\left(\frac{p}{2}+1\right) \log g^{\star}-\log h^{\star}-\log a_0^{\star}-\log k_0^{\star}, \qquad (7)
\end{aligned}
$$

其中所有量的定义见补充材料第 D.2 节。该方法需要计算 $\boldsymbol{\Sigma}_\beta^{\star} \mathbf{X}$ $\mathbb{E}_w[w]$ 以及 $\operatorname{diag}\left(\boldsymbol{\Sigma}_\beta^{\star}\right)$，可通过应用补充材料第 E.1节中的策略高效地求得。

D 网络附录 D

这里报告了用于计算补充材料 $\mathcal{L}$ 中（6）和（7）下界的各项组成。以下结果依赖于性质 $\mathbb{E}_x[\mathbf{x}^{\intercal} \mathbf{A} \mathbf{x}]=\mathbb{E}_x(\mathbf{x}) \mathbf{A} \mathbb{E}_x(\mathbf{x})+tr(\mathbf{A} \mathbf{V})$，其中 $\mathbf{V}$ 是随机变量 $\mathbf{x}$的相关矩阵。

D.1 线性回归

变分下界 $\mathcal{L}$ （6）通过以下各项计算得出：

$$\mathbb{E}_{q^\star}\left[\log p\left(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}\right)\right] \propto -\frac{n}{2}\mathbb{E}_{\sigma^2}\left[\log \sigma^2\right] - \frac{1}{2}\mathbb{E}_{\beta\cdot\sigma^2}\left[\frac{\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2}{\sigma^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\beta} \mid \sigma^2, \tau^2, \lambda_0^2, \boldsymbol{\lambda}\right)\right] \propto -\frac{p+1}{2}\mathbb{E}_{\sigma^2}\left[\log \sigma^2\right] - \frac{p+1}{2}\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \frac{1}{2}\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \sum_{j=1}^{p}\mathbb{E}_{\lambda}\left[\log \lambda_j\right] - \frac{1}{2}\sum_{j=0}^{p}\mathbb{E}_{\beta_0\cdot\beta\cdot\lambda_0^2\cdot\lambda\cdot\tau^2\cdot\sigma^2}\left[\frac{\beta_j^2}{\sigma^2\tau^2\lambda_j^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\lambda_0^2 \mid \psi_0\right)\right] \propto -\frac{1}{2}\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \frac{3}{2}\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \mathbb{E}_{\lambda_0^2\cdot\psi_0}\left[\frac{1}{\psi_0\lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\psi_0\right)\right] \propto -\frac{3}{2}\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \mathbb{E}_{\psi_0}\left[\psi_0^{-1}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\gamma}, \boldsymbol{\varphi}^2\right)\right] \propto \sum_{j=1}^{p}\left(-\log k_j - \frac{1}{2}\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda\cdot\gamma\cdot\varphi^2}\left[\frac{\left(\lambda_j - \mathbf{z}_j^\intercal\boldsymbol{\gamma}\right)^2}{\varphi_j^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\varphi}^2\right)\right] \propto -\frac{1}{2}\sum_{j=1}^{p}\left(3\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] + \mathbb{E}_{\varphi^2}\left[\varphi_j^{-2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\gamma} \mid \boldsymbol{\kappa}^2\right)\right] \propto -\frac{1}{2}\sum_{d=1}^{D}\left(m_d\mathbb{E}_{\kappa^2}\left[\log \kappa_d^2\right] + \mathbb{E}_{\gamma\cdot\kappa^2}\left[\frac{\boldsymbol{\gamma}_d^\intercal\boldsymbol{\gamma}_d}{\kappa_d^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\kappa}^2\right)\right] \propto -\sum_{d=1}^{D}\left((a_d+1)\mathbb{E}_{\kappa^2}\left[\log \kappa_d^2\right] + b_d\mathbb{E}_{\kappa^2}\left[\kappa_d^{-2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\tau^2 \mid \zeta\right)\right] \propto -\frac{1}{2}\mathbb{E}_{\zeta}\left[\log \zeta\right] - \frac{3}{2}\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \mathbb{E}_{\tau^2\cdot\zeta}\left[\frac{1}{\zeta\tau^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\zeta\right)\right] \propto -\frac{3}{2}\mathbb{E}_{\zeta}\left[\log \zeta\right] - \mathbb{E}_{\zeta}\left[\zeta^{-1}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\sigma^2\right)\right] \propto -\left(v+1\right)\mathbb{E}_{\sigma^2}\left[\log \sigma^2\right] - q\mathbb{E}_{\sigma^2}\left[\sigma^{-2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\beta}\right)\right] \propto -\frac{p+1}{2}\mathbb{E}_{\sigma^2}\left[\log \sigma^2\right] - \frac{1}{2}\log\left|\boldsymbol{\Sigma}_\beta^\star\right| - \frac{p+1}{2}$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\lambda_0^2\right)\right] \propto \log\left(a_0^\star\right) - 2\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \mathbb{E}_{\lambda_0^2\cdot\psi_0}\left[\frac{1}{\psi_0\lambda_0^2}\right] - \frac{1}{2}\mathbb{E}_{\beta_0\cdot\lambda_0^2\cdot\tau^2\cdot\sigma^2}\left[\frac{\beta_0^2}{\sigma^2\tau^2\lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\psi_0\right)\right] \propto \log k_0^\star - 2\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \mathbb{E}_{\psi_0}\left[\psi_0^{-1}\right] - \mathbb{E}_{\lambda_0^2\cdot\psi_0}\left[\frac{1}{\psi_0\lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\lambda}\right)\right] \propto \sum_{j=1}^{p}\left(-\log s_j - \mathbb{E}_{\lambda}\left[\log \lambda_j\right] - \frac{1}{2}\mathbb{E}_{\beta\cdot\lambda\cdot\sigma^2\cdot\tau^2}\left[\frac{\beta_j^2}{\sigma^2\tau^2\lambda_j^2}\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda\cdot\varphi^2}\left[\frac{\lambda_j^2}{\varphi_j^2}\right] + \frac{1}{s_0^2}\mathbf{z}_j^{\intercal}\mathbb{E}_{\lambda\cdot\gamma\cdot\varphi^2}\left[\frac{\boldsymbol{\gamma}\lambda_j}{\varphi_j^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\varphi}^2\right)\right] \propto \sum_{j=1}^{p}\left(\log d_j^\star - 2\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] - \frac{1}{2}\mathbb{E}_{\varphi^2}\left[\varphi_j^{-2}\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda\cdot\gamma\cdot\varphi^2}\left[\frac{\left(\lambda_j - \mathbf{z}_j^{\intercal}\boldsymbol{\gamma}\right)^2}{\varphi_j^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\gamma}\right)\right] \propto -\frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\gamma}^\star\right| - \frac{M}{2s_0^2}$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\kappa^2\right)\right] \propto \sum_{d=1}^{D}\left(\left(a_d + \frac{m_d}{2}\right)\log f_d^\star - b_d\mathbb{E}_{\kappa^2}\left[\kappa_d^{-2}\right] - \left(a_d + \frac{m_d}{2} + 1\right)\mathbb{E}_{\kappa^2}\left[\log \kappa_d^2\right] - \frac{1}{2}\mathbb{E}_{\gamma\cdot\kappa^2}\left[\frac{\boldsymbol{\gamma}_d^{\intercal}\boldsymbol{\gamma}_d}{\kappa_d^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\tau^2\right)\right] \propto \left(\frac{p}{2} + 1\right)\log g^\star - \left(\frac{p}{2} + 2\right)\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \mathbb{E}_{\tau^2\cdot\zeta}\left[\frac{1}{\zeta\tau^2}\right] - \frac{1}{2}\sum_{j=0}^{p}\mathbb{E}_{\beta_0\cdot\beta\cdot\lambda_0^2\cdot\lambda\cdot\tau^2\cdot\sigma^2}\left[\frac{\beta_j^2}{\sigma^2\tau^2\lambda_j^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\text{对数 } q^\star\left(\zeta\right)\right] \propto \text{对数 } h^\star - 2\mathbb{E}_{\zeta}\left[\text{对数 } \zeta\right] - \mathbb{E}_{\zeta}\left[\zeta^{-1}\right] - \mathbb{E}_{\tau^2\cdot\zeta}\left[\frac{1}{\zeta\tau^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\text{对数 } q^\star\left(\sigma^2\right)\right] \propto \left(v + \frac{n+p+1}{2}\right)\text{对数 } l^\star - q\mathbb{E}_{\sigma^2}\left[\sigma^{-2}\right] - \left(v + \frac{n+p+3}{2}\right)\mathbb{E}_{\sigma^2}\left[\text{对数 } \sigma^2\right] - \frac{1}{2}\mathbb{E}_{\beta\cdot\sigma^2}\left[\frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{\sigma^2}\right] - \frac{1}{2}\sum_{j=0}^{p}\mathbb{E}_{\beta_0\cdot\beta\cdot\lambda_0^2\cdot\lambda\cdot\sigma^2\cdot\tau^2}\left[\frac{\beta_j^2}{\sigma^2\tau^2\lambda_j^2}\right]$$

D.2 Probit 回归

变分下界 $\mathcal{L}$ （7）由补充材料中的以下部分计算得出:

$$\mathbb{E}_{q^\star}\left[\log p\left(\mathbf{y} \mid \mathbf{w}\right)\right] \propto \sum_{i=1}^{n}\left(\mathbb{E}_w\left[y_i \log(\mathbb{I}_{(w_i>0)}\right] + \mathbb{E}_w\left[(1-y_i)\log(\mathbb{I}_{(w_i<0)}\right]\right) = 0$$

$$\mathbb{E}_{q^\star}\left[\log p\left(\mathbf{w} \mid \boldsymbol{\beta}\right)\right] \propto -\frac{1}{2}\sum_{i=1}^{n}\mathbb{E}_w\left[\|w_i - \mathbf{x}_i^\intercal\boldsymbol{\beta}\|_2^2\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\beta} \mid \sigma^2, \tau^2, \lambda_0^2, \boldsymbol{\lambda}\right)\right] \propto -\frac{p+1}{2}\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \frac{1}{2}\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \sum_{j=1}^{p}\mathbb{E}_{\lambda}\left[\log \lambda_j\right] - \frac{1}{2}\sum_{j=0}^{p}\mathbb{E}_{\beta_0\cdot\beta\cdot\lambda_0^2\cdot\lambda\cdot\tau^2}\left[\frac{\beta_j^2}{\tau^2\lambda_j^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\lambda_0^2 \mid \psi_0\right)\right] \propto -\frac{1}{2}\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \frac{3}{2}\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \mathbb{E}_{\lambda_0^2\cdot\psi_0}\left[\frac{1}{\psi_0\lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\psi_0\right)\right] \propto -\frac{3}{2}\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \mathbb{E}_{\psi_0}\left[\psi_0^{-1}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\gamma}, \boldsymbol{\varphi}^2\right)\right] \propto \sum_{j=1}^{p}\left(-\log k_j - \frac{1}{2}\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda\cdot\gamma\cdot\varphi^2}\left[\frac{\left(\lambda_j - \mathbf{z}_j^\intercal\boldsymbol{\gamma}\right)^2}{\varphi_j^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\varphi}^2\right)\right] \propto -\frac{1}{2}\sum_{j=1}^{p}\left(3\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] + \mathbb{E}_{\varphi^2}\left[\varphi_j^{-2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\gamma} \mid \boldsymbol{\kappa}^2\right)\right] \propto -\frac{1}{2}\sum_{d=1}^{D}\left(m_d\mathbb{E}_{\kappa}\left[\log \kappa_d^2\right] + \mathbb{E}_{\gamma\cdot\kappa}\left[\frac{\boldsymbol{\gamma}_d^\intercal\boldsymbol{\gamma}_d}{\kappa_d^2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\boldsymbol{\kappa}^2\right)\right] \propto \sum_{d=1}^{D}\left(-(a_d+1)\mathbb{E}_{\kappa^2}\left[\log \kappa_d^2\right] - b_d\mathbb{E}_{\kappa^2}\left[\kappa_d^{-2}\right]\right)$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\tau^2 \mid \zeta\right)\right] \propto -\frac{1}{2}\mathbb{E}_{\zeta}\left[\log \zeta\right] - \frac{3}{2}\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \mathbb{E}_{\tau^2\cdot\zeta}\left[\frac{1}{\zeta\tau^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log \pi\left(\zeta\right)\right] \propto -\frac{3}{2}\mathbb{E}_{\zeta}\left[\log \zeta\right] - \mathbb{E}_{\zeta}\left[\zeta^{-1}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\mathbf{w}\right)\right] \propto \sum_{i=1}^{n}\left(-\frac{1}{2}\mathbb{E}_w\left[\|w_i - \mathbf{x}_i^\intercal\boldsymbol{\beta}\|_2^2\right] - y_i\log\left(1-\Phi_i\right) - (1-y_i)\log\left(\Phi_i\right)\right)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\beta}\right)\right] \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_\beta^\star| - \frac{p+1}{2}$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\lambda_0^2\right)\right] \propto \log\left(a_0^\star\right) - 2\mathbb{E}_{\lambda_0^2}\left[\log \lambda_0^2\right] - \mathbb{E}_{\lambda_0^2\cdot\psi_0}\left[\frac{1}{\psi_0\lambda_0^2}\right] - \frac{1}{2}\mathbb{E}_{\beta_0\cdot\tau^2\cdot\lambda_0^2}\left[\frac{\beta_0^2}{\tau^2\lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\psi_0\right)\right] \propto \log k_0^\star - 2\mathbb{E}_{\psi_0}\left[\log \psi_0\right] - \mathbb{E}_{\psi_0}\left[\psi_0^{-1}\right] - \mathbb{E}_{\lambda_0^2 \cdot \psi_0}\left[\frac{1}{\psi_0 \lambda_0^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\lambda}\right)\right] \propto \sum_{j=1}^{p}\Bigg(- \log s_j - \mathbb{E}_{\lambda}\left[\log \lambda_j\right] - \frac{1}{2}\mathbb{E}_{\beta \cdot \tau^2 \cdot \lambda}\left[\frac{\beta_j^2}{\tau^2 \lambda_j^2}\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda \cdot \varphi^2}\left[\frac{\lambda_j^2}{\varphi_j^2}\right] + \frac{1}{s_0^2}\mathbf{z}_j^\intercal \mathbb{E}_{\lambda \cdot \gamma \cdot \varphi^2}\left[\frac{\boldsymbol{\gamma}\lambda_j}{\varphi_j^2}\right]\Bigg)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\varphi}^2\right)\right] \propto \sum_{j=1}^{p}\Bigg(\log d_j^\star - 2\mathbb{E}_{\varphi^2}\left[\log \varphi_j^2\right] - \frac{1}{2}\mathbb{E}_{\varphi^2}\left[\varphi_j^{-2}\right] - \frac{1}{2s_0^2}\mathbb{E}_{\lambda \cdot \gamma \cdot \varphi^2}\left[\frac{\left(\lambda_j - \mathbf{z}_j^\intercal \boldsymbol{\gamma}\right)^2}{\varphi_j^2}\right]\Bigg)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\boldsymbol{\gamma}\right)\right] \propto -\frac{1}{2}\log\left|\boldsymbol{\Sigma}_\gamma^\star\right| - \frac{M}{2s_0^2}$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\kappa^2\right)\right] \propto \sum_{d=1}^{D}\Bigg(\left(a_d + \frac{m_d}{2}\right)\log f_d^\star - (a_d + \frac{m_d}{2} + 1)\mathbb{E}_{\kappa^2}\left[\log \kappa_d^2\right] - b_d \mathbb{E}_{\kappa^2}\left[\kappa_d^{-2}\right] - \frac{1}{2}\mathbb{E}_{\gamma \cdot \kappa^2}\left[\frac{\boldsymbol{\gamma}_d^\intercal \boldsymbol{\gamma}_d}{\kappa_d^2}\right]\Bigg)$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\tau^2\right)\right] \propto \left(\frac{p}{2} + 1\right)\log\left(g^\star\right) - \left(\frac{p}{2} + 2\right)\mathbb{E}_{\tau^2}\left[\log \tau^2\right] - \mathbb{E}_{\tau^2}\left[\frac{1}{\zeta\tau^2}\right] - \frac{1}{2}\sum_{j=0}^{p}\mathbb{E}_{\beta_0 \cdot \beta \cdot \lambda_0^2 \cdot \lambda \cdot \tau^2}\left[\frac{\beta_j^2}{\tau^2 \lambda_j^2}\right]$$

$$\mathbb{E}_{q^\star}\left[\log q^\star\left(\zeta\right)\right] \propto \log h^\star - 2\mathbb{E}_{\zeta}\left[\log \zeta\right] - \mathbb{E}_{\zeta}\left[\zeta^{-1}\right] - \mathbb{E}_{\zeta}\left[\frac{1}{\zeta\tau^2}\right],$$

where $\Phi_i = \Phi\left(-\mathbf{x}_i^\intercal \mathbb{E}_\beta\left[\boldsymbol{\beta}\right]\right)$.

E 网络附录 E

E.1 变分算法的计算方法

变分算法的优点是不需要显式求逆协方差矩阵 $\boldsymbol{\Sigma}_{\beta}^{\star}=(\mathbf{X}\,\mathbf{X}\,+\boldsymbol{\Delta})^{-1}$，其中 $\boldsymbol{\Delta}=\mathbb{E}_{\lambda\cdot\tau^{2}}\left[\tau^{-2}\Lambda^{-2}\right]$ 为对角矩阵；当协变量数量 p 增加时，显式求逆在计算上会变得不可行。所需的量包括 $\boldsymbol{\Sigma}_{\beta}^{\star}$的对角元素、后验均值 $\boldsymbol{\mu}_{\beta}^{\star}=\boldsymbol{\Sigma}_{\beta}^{\star}(\mathbf{X}\,\mathbf{y})$ 和 $tr\left(\mathbf{X}\,\mathbf{X}\boldsymbol{\Sigma}_{\beta}^{\star}\right)$。利用伍德伯里恒等式，可以高效地计算所需的量：

$$\begin{aligned}
\operatorname{diag}\left(\boldsymbol{\Sigma}_{\beta}^{\star}\right) &= \operatorname{diag}\left(\boldsymbol{\Delta}^{-1}\right)-\operatorname{diag}\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\left(\mathbf{I}_{n}+\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)^{-1}\mathbf{X}\boldsymbol{\Delta}^{-1}\right)\\
&= \operatorname{diag}\left(\boldsymbol{\Delta}^{-1}\right)-\left[\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)\left(\mathbf{I}_{n}+\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)^{-1}\circ\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)\right]\cdot\mathbf{1}_{n\times 1}, \quad (8)\\
\boldsymbol{\mu}_{\beta}^{\star} &= \left(\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{X}^{\intercal}\right)\mathbf{y}\\
&= \left[\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)-\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)\left(\mathbf{I}_{n}+\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)^{-1}\left(\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)\right]\mathbf{y}, \quad (9)\\
tr\left(\mathbf{X}\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{X}^{\intercal}\right) &= \sum_{i=1}^{n}\mathbf{x}_{i}^{\intercal}\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{x}_{i}=\sum_{i=1}^{n}\mathbf{x}_{i}^{\intercal}\left(\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{x}_{i}\right), \quad (10)
\end{aligned}$$

其中 $\circ$ 表示哈达玛积，$\mathbf{I}_{n}$ 是 $n\times n$ 单位矩阵，且 $\mathbf{1}_{n\times 1}$表示 n维全 1 向量。在式（10）中，量 $\boldsymbol{\Sigma}_{\beta}^{\star}\mathbf{x}_{i}$ 在计算向量 $\boldsymbol{\mu}_{\beta}^{\star}$ 时已经求得，而式（8）—（10）中的所有矩阵乘积均可通过 $\mathcal{O}\left(n^{2}p\right)$ 次运算求得，其复杂度关于 p呈线性。请注意，在 Probit 回归中不需要最后一个量（10），因此后验均值可以高效地计算为：

$$\boldsymbol{\mu}_{\beta}^{\star}=\boldsymbol{\Delta}^{-1}\left(\mathbf{X}^{\intercal}\mathbb{E}_{w}\left[\mathbf{w}\right]\right)-\left(\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)\left(\mathbf{I}_{n}+\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right)^{-1}\left(\mathbf{X}\boldsymbol{\Delta}^{-1}\right)\left(\mathbf{X}^{\intercal}\mathbb{E}_{w}\left[\mathbf{w}\right]\right),$$

这只涉及矩阵-向量乘积，进一步提高了算法的计算效率。最后，可以高效地计算行列式：

$$\left|\boldsymbol{\Sigma}_{\beta}^{\star}\right|=\left|\boldsymbol{\Delta}\right|/\left|\mathbf{I}_{n}+\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{\intercal}\right|.$$

F 网络附录 F

F.1 模拟研究的详细信息

对于所有考虑的病例，我们采用以下模拟方案。设 $\boldsymbol{\beta}^{0}$ 为真实的（$p+1$维回归参数向量。我们将真实非零系数的数量设为 p_{0} （不包括截距）。设计矩阵的各项独立采样自 $x_{ij}\sim\mathcal{N}(0,1),\ i=1,\ldots,n,\ j=1,\ldots,p$，而响应变量 $\mathbf{y}$ 和（$p+1$维真实回归系数向量则按照 Johnson(2013中方案的一个修改版本进行采样。具体而言，

$$y_i = \mathbf{x}_i^\intercal \boldsymbol{\beta}^0 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1), \quad i = 1, \ldots, n,$$
$$\beta_j^0 = \begin{cases} v_0 |t| & \text{if } j = 0, \\ (-1)^u \left(v^2 \log(n)/\sqrt{n} + v|t| \right) & \text{if } j = 1, \ldots, p_0, \\ 0 & \text{otherwise}, \end{cases}$$

其中 $v_0^2 = 0.5$、 $v^2 = 0.75$、 $u \sim \mathcal{B}(0.4)$ 和 $t \sim \mathcal{N}(0,1)$。

F.2 协变量数据模拟

这里说明如何模拟协变量信息，以评估正文5 节中变分近似相对于吉布斯采样器的准确性。此外，我们还提供了两种方法在变量选择方面的曲线下面积图示。

四种程度的协变量信息模拟如下：

$G0$) **无协变量数据** 设置：协变量数据矩阵中仅包含截距，因此$\mathbf{Z} = \mathbf{1}_p$;

$G1$) **非信息性** 设置：通过随机选择 30 个回归变量，在模型中加入一个二元协变量数据源，因此协变量数据矩阵$\mathbf{Z}$ 由二元向量$\mathbf{z} \in \{0,1\}^p$构成，其中 $z_j = 1$ 若选中第 j 个变量则为 $z_j = 0$ ，否则为 $z_j = 0$ ；

$G2$) **信息性** 设置：通过随机选择 20 个真实非零回归变量和 10 个真实零回归变量，在模型中加入一个二元协变量数据源，因此协变量数据矩阵$\mathbf{Z}$ 由二元向量 $\mathbf{z} \in \{0,1\}^p$构成，其中 $z_j = 1$ 若选中第 j 个变量则为 $z_j = 0$ ，否则为 $z_j = 0$ ；

$G3$) **完美协变量数据信息** 设定：协变量数据矩阵 $\mathbf{Z}$由二元向量 $\mathbf{z} \in \{0,1\}^p$构建，其中 $z_j = 1$ 若 $\beta_j^0 \neq 0$， $z_j = 0$ 否则。

G 网络附录 G

G.1 Web算法 1

算法 1： 信息性马蹄回归的吉布斯采样器

输入：B，$bn \in \mathbb{N}$，$\mathbf{y} \in \mathbb{R}^n$，$\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$，$\mathbf{Z}_1$，…，$\mathbf{Z}_D \in \mathbb{R}^{p\times m_d}$，$\mathbf{a}$，$\mathbf{b} \in \mathbb{R}_+^D$，$s_0^2 \in \mathbb{R}_+$；

// 设置 $\lambda_0 = 1$，$\boldsymbol{\lambda} = \mathbf{1}_p$、$\boldsymbol{\gamma} = \mathbf{0}_M$和 σ^2，$\tau^2 = 1$并采样参数 $\boldsymbol{\varphi}$、$\boldsymbol{\kappa}^2$ 和 ζ 从其先验分布中

对于 $b = 1$ 到 B **做**

样本 $\boldsymbol{\beta} \,|\, \mathbf{y}$，$\mathbf{X}$ σ^2，τ^2，λ_0^2，$\boldsymbol{\lambda} \sim \mathcal{N}_{p+1}\left(\boldsymbol{\Sigma}_\beta^\star \mathbf{X}\mathbf{y}\ \sigma^2\boldsymbol{\Sigma}_\beta^\star\right)$，其中 $\boldsymbol{\Sigma}_\beta^\star = \left(\mathbf{X}\mathbf{X} + \tau^{-2}\boldsymbol{\Lambda}^{-2}\right)^{-1}$；

样本 $\lambda_0^2 \,|\, \beta_0$，$\tau^2$，$\sigma^2 \sim \mathcal{IG}\left(1, \frac{1}{\psi_0} + \frac{\beta_0^2}{2\sigma^2\tau^2}\right)$；

样本 $\psi_0 \,|\, \lambda_0^2 \sim \mathcal{IG}\left(1, 1 + \frac{1}{\lambda_0^2}\right)$；

对于 $j = 1$ 至 p，**执行**

抽取样本 $\lambda_j \,|\, \mathbf{Z}\ \beta_j, \sigma^2, \tau^2, \boldsymbol{\gamma}, \varphi_j^2 \propto \lambda_j^{-1} e^{-\frac{\beta_j^2}{2\sigma^2\tau^2\lambda_j^2} - \frac{\lambda_j^2}{2s_0^2\varphi_j^2} + \frac{\mu_j\lambda_j}{s_0^2\varphi_j^2}} \cdot \mathbb{I}_{(\lambda_j>0)}$ 接下来补充材料第 B.1 节中的程序；

抽取样本 $\varphi_j^2 \,|\, \mathbf{Z}\ \lambda_j, \boldsymbol{\gamma} \sim \mathcal{IG}\left(1, \frac{1}{2} + \frac{(\lambda_j-\mu_j)^2}{2s_0^2}\right)$；

end

样本 $\boldsymbol{\gamma} \,|\, \mathbf{Z}\ \boldsymbol{\lambda}$，$\boldsymbol{\varphi}$，$\boldsymbol{\kappa}^2 \sim \mathcal{N}_M\left(\boldsymbol{\Sigma}_\gamma^\star\left(\mathbf{Z}^\intercal\boldsymbol{\Phi}^{-2}\boldsymbol{\lambda}\right),\ s_0^2\boldsymbol{\Sigma}_\gamma^\star\right)$，其中 $\boldsymbol{\Sigma}_\gamma^\star = \left(\mathbf{Z}^\intercal\boldsymbol{\Phi}^{-2}\mathbf{Z} + s_0^2\mathbf{D}_\kappa^{-1}\right)^{-1}$和 $\mathbf{D}_\kappa$ = 对角矩阵 $\left(\kappa_1^2\mathbf{1}_{m_1}, \ldots, \kappa_D^2\mathbf{1}_{m_D}\right)$；

对于 $d = 1$ 至 D **执行**

样本 $\kappa_d^2 \,|\, \boldsymbol{\gamma}_d \sim \mathcal{IG}\left(a_d + \frac{m_d}{2}, b_d + \frac{\boldsymbol{\gamma}_d^\intercal\boldsymbol{\gamma}_d}{2}\right)$；

end

样本 $\tau^2 \,|\, \boldsymbol{\beta}$，$\sigma^2$，$\lambda_0^2$，$\boldsymbol{\lambda}$，$\zeta \sim \mathcal{IG}\left(\frac{p}{2} + 1, \frac{1}{\zeta} + \frac{\beta_0^2}{2\sigma^2\lambda_0^2} + \frac{1}{2\sigma^2}\sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right)$；

样本 $\zeta \,|\, \tau^2 \sim \mathcal{IG}\left(1, 1 + \frac{1}{\tau^2}\right)$；

样本 $\sigma^2 \,|\, \mathbf{y}$，$\mathbf{X}$ $\boldsymbol{\beta}$，τ^2，λ_0^2，$\boldsymbol{\lambda}^2 \sim \mathcal{IG}\left(v + \frac{n+p+1}{2}, q + \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \frac{\beta_0^2}{2\tau^2\lambda_0^2} + \frac{1}{2\tau^2}\sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right)$；

如果 $b > bn$ **那么**

保存 $\boldsymbol{\beta}$, σ^2, τ^2, λ_0^2, $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$ 和 $\boldsymbol{\kappa}^2$

end

结束

返回 已保存的值；

G.2 Web算法 2

算法 2： 针对信息性 Horseshoe 回归的变分贝叶斯近似回归

输入：$\mathbf{y} \in \mathbb{R}^n$，$\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$，$\mathbf{Z}_1$，$\ldots$，$\mathbf{Z}_D \in \mathbb{R}^{p\times m_d}$，$v$，$q \in \mathbb{R}^+$，$\mathbf{a}$，$\mathbf{b} \in \mathbb{R}_+^D$，$s_0^2 \in \mathbb{R}_+$；

// 设置 $b = 1$，$\epsilon = 10^{-3}$，$\mathcal{L}^{(0)} = -\infty$ 并初始化所有所需的矩

当 $\mathcal{L}^{(b)} - \mathcal{L}^{(b-1)} > \epsilon$ **时执行**

- 更新参数 $\boldsymbol{\mu}_\beta^\star$，并计算对角矩阵 $\left(\boldsymbol{\Sigma}_\beta^\star\right)$、$|\boldsymbol{\Sigma}_\beta^\star|$ 和 $\mathbf{X}\boldsymbol{\Sigma}_\beta^\star\mathbf{X}^\intercal$ 如补充材料的第 E.1 节所述；
- 更新参数 $a_0^\star$ 和 $k_0^\star$；
- 更新参数 $a_j^\star$、$b_j^\star$、$c_j^\star$ 和 $d_j^\star$，并计算归一化常数 s_j，$\mathbb{E}_\lambda[\lambda_j]$、$\mathbb{E}_\lambda\left[\lambda_j^2\right]$ 和 $\mathbb{E}_\lambda\left[\lambda_j^{-2}\right]$ 通过数值积分，对 $j = 1, \ldots, p$；
- 更新参数 $\boldsymbol{\mu}^\star{}_\gamma$ 和 $\boldsymbol{\Sigma}^\star{}_\gamma$；
- 更新参数 $e_d^\star$和 $f_d^\star$，对 $d = 1, \ldots, D$；
- 更新参数 $g^\star$ 和 $h^\star$；
- 更新参数 $l^\star$；
- 计算 $\mathcal{L}^{(b)}$ 并设置 $b = b + 1$；

结束

返回 $\boldsymbol{\mu}_\beta^\star$, $\boldsymbol{\Sigma}_\beta^\star$, $a_0^\star$, $\mathbf{a}^\star$, $\mathbf{b}^\star$, $\mathbf{c}^\star$, $\boldsymbol{\mu}_\gamma^\star$, $\boldsymbol{\Sigma}_\gamma^\star$, $\mathbf{e}^\star$, $\mathbf{f}^\star$, $g^\star$ 和 $l^\star$；

G.3 Web算法 3

算法 3： Probit 信息性
马蹄回归的变分贝叶斯近似

输入：$\mathbf{y} \in \{0,1\}^n$，$\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$，$\mathbf{Z}_1$、…、$\mathbf{Z}_D \in \mathbb{R}^{p\times m_d}$，$\mathbf{a}$，$\mathbf{b} \in \mathbb{R}_+^D$，$s_0^2 \in \mathbb{R}_+$；
// 设置 $b=1$，$\epsilon = 10^{-3}$，$\mathcal{L}^{(0)} = -\infty$ 并初始化所有所需的矩
当 $\mathcal{L}^{(b)} - \mathcal{L}^{(b-1)} > \epsilon$ **满足时**
  更新参数 $\mu_i^\star$，对于 $i = 1, \ldots, n$；
  更新参数 $\boldsymbol{\mu}_\beta^\star$ 并计算 $\operatorname{diag}\left(\boldsymbol{\Sigma}_\beta^\star\right)$ 和 $|\boldsymbol{\Sigma}_\beta^\star|$，如章节E.1 补充材料；
  更新参数 $a_0^\star$ 和 $k_0^\star$；
  更新参数 $a_j^\star$、$b_j^\star$、$c_j^\star$ 和 $d_j^\star$，并计算归一化常数 s_j，通过数值积分计算 $\mathbb{E}_\lambda[\lambda_j]$、$\mathbb{E}_\lambda\left[\lambda_j^2\right]$ 和 $\mathbb{E}_\lambda\left[\lambda_j^{-2}\right]$，对于 $j = 1$、……、p；
  更新参数 $\boldsymbol{\mu}^\star$ $_\gamma$ 以及 $\boldsymbol{\Sigma}_\gamma^\star$；
  更新参数 $e_d^\star$ 和 $f_d^\star$，对于 $d = 1$、……、D；
  更新参数 $g^\star$ 和 $h^\star$；
  计算 $\mathcal{L}^{(b)}$ 并设置 $b = b+1$；
结束
返回 $\boldsymbol{\mu}_\beta^\star$, $a_0^\star$, $\mathbf{a}^\star$, $\mathbf{b}^\star$, $\mathbf{c}^\star$, $\boldsymbol{\mu}_\gamma^\star$, $\boldsymbol{\Sigma}_\gamma^\star$, $\mathbf{e}^\star$, $\mathbf{f}^\star$, $g^\star$；

H 网络附录 H

H.1 网络图 1

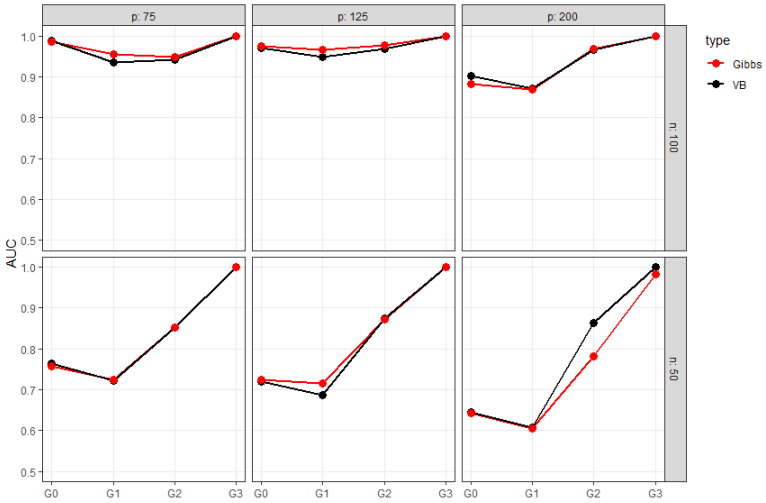


图5：使用吉布斯采样器和变分算法进行变量选择； AUC是对每项实验的10次重复实验取平均所得。

H.2 网络图 2

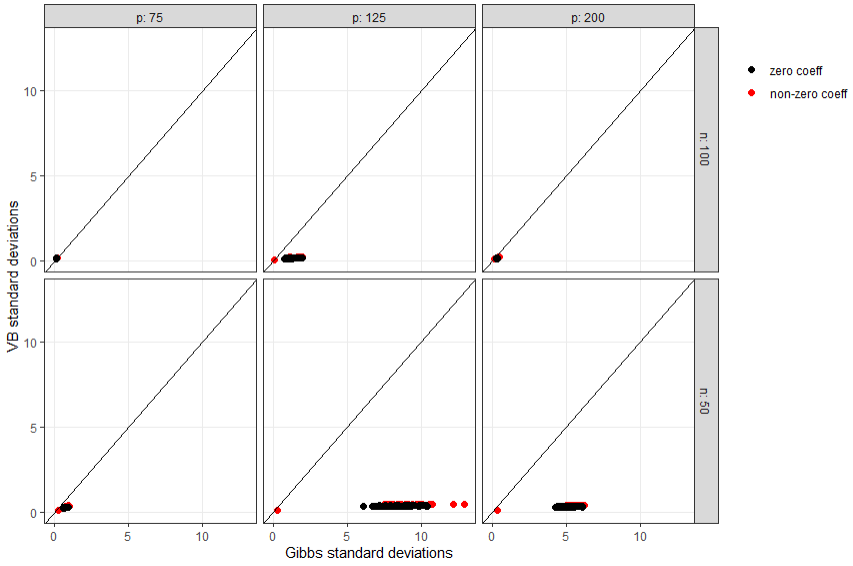


图 6：使用吉布斯采样器和变分算法估计的回归系数标准差 β；结果基于每个实验的 10 次重复实验进行评估。数据已按不同类型的协变量数据汇总。