arXiv:2611.28413v1 [stat.CO] 20 Nov 2026

# 具有多个重尾随机效应分量和未知尾部参数的线性混合模型的吉布斯采样器的几何遍历性

Rolando de la Cruz* Nathalie Humeniy†

2026年11月24日

**摘要**

具有重尾随机效应的线性混合模型广泛用于纵向数据的稳健贝叶斯分析。我们研究这类模型的三块吉布斯采样器，其中随机效应被划分为多个分量，例如随机截距和斜率，每个分量都有自己的尺度以及自己的正态尺度混合；尾部参数的先验分布定义在有限网格上，并与潜在权重联合抽样；随机效应设计矩阵可能秩亏，例如某些受试者仅贡献一次观测。我们给出了几何遍历性的充分条件，将秩亏归因于导致秩亏的分量；并证明，在精度参数的伽马先验满足一定条件时，对于带有个体权重的 Student-t 分布随机效应和 slash随机效应、任意尾部参数网格以及任意设计矩阵，这些条件均成立。该结果扩展到坐标共享同一权重的随机效应向量，以及具有已知形状矩阵的个体特异性设计。该结果为回归系数、对数尺度和尾部参数后验概率的蒙特卡罗标准误提供了理论依据。模拟结果表明，尺度和尾部参数的混合速度可能远低于回归系数；对 beta-HCG 数据的分析则展示了这些方法的应用。

**关键词：** 漂移条件；重尾随机效应；多个方差成分；秩亏设计矩阵；Student-$t$ 自由度；马尔可夫链中心极限定理。

## 1 引言

具有重尾随机效应的线性混合模型广泛用于处理其轨迹偏离总体大多数受试者的情况（Lange等，1989年；Pinheiro 等人，2001年；Rosa等，2003年；Lachos等，2010年）。在贝叶斯分析中，随机效应通常表示为正态分布的尺度混合：给定潜在精度权重时，每个随机效应服从正态分布，而这些权重服从某种混合分布 $H$，例如产生 Student-$t$随机效应的伽马分布。采用条件共轭先验时，这种表示方式可得到一个包含三个分块的吉布斯采样器：方差分量、潜在权重，以及回归系数与随机效应。

纵向研究通常会结合三个特征，使该采样器的收敛更难分析：存在多个尺度、尾部情况不确定，而且每位受试者的观测值很少。这些因素的相互作用正是本文的研究主题。

*工程与科学学院，Universidad Adolfo Ibáñez，Peñalolén，智利；Data Observatory Foundation，Santiago，智利；社会数据科学千禧年核心（SODAS），Santiago，智利。通讯作者：rolando.delacruz@uai.cl

†统计学系，Pontificia Universidad Católica de Chile，Santiago，智利。

首先，随机效应被拆分为 $K$ 多个分量，例如随机截距和随机斜率，或不同受试者组的效应；每个分量都有自己的尺度参数和混合分布，且随机效应设计矩阵 $Z$ 是秩亏的。对于随机截距和斜率，每位仅有单次观测的受试者都会产生 $Z$的两列成比例列；在 $\beta$第5节的 HCG 数据中，173 名女性里有 52 名仅有一次测量，且秩$(Z) = q - 52$。在存在多个尺度和重尾的情况下，潜在权重会与 $Z$ 的零空间逐分量相互作用，因此，必须将秩亏归因于导致秩亏的分量，而不能将其归咎于所有分量。

其次，每个混合分布的尾部参数通常并非已知。可以使用预测准则来选择该参数（de la Cruz 和 Humeniy，2026年），但完整的贝叶斯处理会为其指定先验分布（Fonseca等人，2008年；Villa和Walker，2014年）。连续先验可能需要额外的非共轭更新。若先验分布定义在有限网格上，则可以精确抽取尾部参数，并与权重联合抽样，算法仍然是三块吉布斯采样器。我们研究这一版本。

第三，在稳健随机系数模型中，同一受试者的随机效应向量各坐标通常共享一个权重（Pinheiro 等人，2001年；Lachos等，2010年），这改变了权重与随机效应之间的依赖关系。

几何遍历性与矩条件相结合，可保证马尔可夫链中心极限定理（中心极限定理），进而确保蒙特卡罗标准误有效（Jones，2004年；Flegal等，2008年；Vats等，2019年）。对于正态随机效应，Román和Hobert(2012年，2015年)给出了在具有多个方差分量且任意 $Z$时该采样器具有几何遍历性的条件；随后Abrahamsen和Hobert (2017年)去除了其中一些假设。对于混合分布已知的随机效应，de la Cruz 和 Humeniy (2026年)证明，对于单个尺度参数和任意设计矩阵，以及在 $Z$ 满列秩时按已知分组索引的多个尺度，模型具有几何遍历性。上述结果未涵盖具有多个尺度、未知尾部参数和共享权重的秩亏设计。

我们的主要结果如下。定理1给出了具有 $K$ 分量、在有限网格上具有任意 $Z$ 和尾部参数时满足几何遍历性的充分条件。关于精度参数的伽马先验形状参数的条件，对于每个分量 $k$，都涉及量 $\delta_k = q_k + r_{-k} - r$，其中 $r = \text{rank}(Z)$，而 $r_{-k}$是不含其第 $k$个块的 $Z$ 的秩，因此只有对秩亏缺有贡献的分量受到限制；对于正态随机效应，这些条件与Román和Hobert（2015）的条件形式相同；当 $K > 1$时，它们并不比后者更强，甚至可能严格更弱。关键要素是对投影到重缩放零空间上的投影矩阵对角线的分块和给出的一个界（引理 5）；尽管权重在每一步都会重缩放零空间，该界仍能逐个分量地归因秩亏缺。关于混合分布的条件通过投影到 $Z$的零空间来表述，而命题 1将其转化为具有未知尾部参数的Student-$t$ 和slash随机效应的条件。将de la Cruz 和 Humeniy (2026)的重缩放论证推广到多个分量和未知尾部参数后（推论1和推论2），对于所有常见分布族，关于尾部的限制都可以消除（推论3）。定理 2允许受试者的随机效应向量的各个坐标共享一个权重，其条件结合了共享权重的倾斜分布与零空间的分块几何结构；推论4进一步推出：对于具有已知形状矩阵、未知组特异尺度、尾部参数取自任意有限网格以及任意个体特异性设计的多元Student-$t$ 和slash随机效应，马尔可夫链具有几何遍历性。

对于贝叶斯实践，结果表明，应计算回归系数、尺度参数的对数以及尾部参数的后验概率的蒙特卡罗标准误，以量化各个分量呈现重尾的证据。第 6节说明了为何应分别计算每个量的这些误差：使用共享重尾权重时，尺度

随机效应的尺度和尾部参数的混合速度可能比回归系数慢数倍，而基于多元有效样本量的多元准则并不能确保每个分量都达到规定的蒙特卡罗精度。

本文其余部分安排如下。第2节介绍模型和采样器。第3 节陈述结果，第4 节给出证明。第5 节将结果应用于$\beta$-HCG 数据，第6 节研究模拟数据中共享权重下对尺度参数和尾部参数的探索，第7 节讨论扩展。附录A 介绍如何验证实现，附录B 证明辅助引理。

## 2 模型和吉布斯采样器

### 2.1 模型

设 $y$为一个 $N\times 1$ 响应变量， $X$为一个 $N\times p$ 固定效应设计矩阵， $Z=(Z_1\cdots Z_K)$为一个随机效应设计矩阵，其分块 $Z_k$为 $N\times q_k$；设 $q=\sum_k q_k$，并且对于 $i=1,\ldots\ldots,q$，令 $k(i)$表示坐标 $i$ 所属的分量。每个分量 $k$ 都有一个尺度参数 $\lambda_k$ 和一个取值于有限集$\mathcal{D}_k\subset\mathbb{R}$中的尾部参数$d_k$。对于每个坐标 $i$ 以及每个 $d\in\mathcal{D}^{k(i)}$，令 $H_{i,\,d}$ 为定义在 $\mathbb{R}_+=(0,\ \infty)$上的概率分布。我们考虑

$$Y\mid\beta,u,\lambda_e\sim \mathrm{N}_N(X\beta+Zu,\lambda_e^{-1}I_N), \tag{1}$$

$$u_i\mid\lambda,\eta,d\overset{\text{ind}}{\sim}\mathrm{N}\big(0,(\lambda_{k(i)}\eta_i)^{-1}\big),\qquad \eta_i\mid d\overset{\text{ind}}{\sim}H_{i,d_{k(i)}},\qquad i=1,\ldots,q, \tag{2}$$

$$\beta\sim\mathrm{N}_p(\mu_\beta,\Sigma_\beta),\quad \lambda_e\sim\mathrm{Gamma}(a_e,b_e),\quad \lambda_k\sim\mathrm{Gamma}(a_k,b_k),\quad d_k\sim\pi_k, \tag{3}$$

其中，$\beta$、$\lambda_e$、$\lambda_1$、……、$\lambda_K$和 $d_1$、……、$d_K$在先验下相互独立，$\Sigma_\beta$正定，所有伽马形状参数和率参数均为正，且 $\pi_k$是$\mathsf{D}_k$上的一个概率测度，其中 $\pi_k(d)>0$ 对每个 $d\in\mathsf{D}_k$均成立。这里，Gamma($a$, $b$)的密度正比于 $x^{a-1}e^{-bx}$。已知尾部参数对应于单点集$\mathsf{D}_k$。在大多数应用中，$H_{i,\,d}=H_{k(i),\,d}$ 不随分量内的坐标而变化；我们允许坐标特定的分布律，因为第3.4节的重缩放论证会产生这类分布律。

给定 $\lambda_{k(i)}$ 和 $d_{k(i)}$，$u_i$ 是正态尺度混合。常见的分布族通过精度乘子 $\eta$参数化，包括自由度为 $t$ 的 Student-$t$ 分布、$H_d=\mathrm{Gamma}(d/2,d/2)$；参数为 $\nu$的 $H_\nu=\mathrm{Beta}(\nu,1)$slash分布；污染正态分布、正态广义逆高斯分布和方差-伽马分布（de la Cruz 和 Humeniy，2026年，表1）。后验分布是适定的，因为似然被$(\lambda_e/2\pi)^{N/2}$所界定，该函数对 $\lambda_e$的伽马先验可积，且所有其他先验分布均适定。

W写出 $\lambda=(\lambda_e,\lambda_1,\ldots,\lambda_K)$，$d=(d_1,\ldots,d_K)$，$\theta=(\beta^T,u^T)^T$，$W=(X\ Z)$和

$$D_\eta=\mathrm{diag}(\eta_1,\ldots,\eta_q),\qquad \Lambda_{\lambda,\eta}=\mathrm{diag}(\lambda_{k(1)}\eta_1,\ldots,\lambda_{k(q)}\eta_q).$$

### 2.2 全条件分布

对于定义在 $\mathbb{R}_+$ 上、具有 $m_{1/2}(H)=\int\eta^{1/2}H(d\eta)<\infty$ 和 $s\geq 0$ 的概率 $H$，令

$$\mathcal{Z}_H(s)=\int_0^\infty \eta^{1/2}e^{-s\eta}H(d\eta),\qquad H_s(d\eta)=\frac{\eta^{1/2}e^{-s\eta}H(d\eta)}{\mathcal{Z}_H(s)}. \tag{4}$$

收集联合后验密度中涉及各分块的因子，可得以下完全条件分布。

**(i)** 给定 $(d,\eta,\theta)$，$\lambda$的各分量相互独立，其中 $\lambda_e\sim$服从伽马分布$\big(a_e+N/2,\ b_e+\|y-W\theta\|^2/2\big)$，且 $\lambda_k\sim$服从伽马分布$\big(a_k+q_k/2,\ b_k+\sum_{i:k(i)=k}\eta_i u_i^2/2\big)$。该分布律不依赖于 $d$。

**算法1** 模型（1）–（3）的分块吉布斯采样器。

**要求：** 当前状态 $(\eta, \theta)$，其中 $\theta = (\beta, u)$

1：分别从各自的伽马全条件分布中抽取 $\lambda_e$ 和 $\lambda_1, \ldots, \lambda_K$。2：对于 $k = 1, \ldots, K$：从 (5) 中抽取 $d_k$，然后对满足 $k(i) = k$ 的 $i$，抽取 $\eta_i \sim (H_{i,d_k})_{s_i}$。3：抽取 $\theta \sim \mathrm{N}_{p+q}(\mathrm{E}(\theta \mid \lambda, \eta)\, P_{\lambda}^{-1}{}_{,\eta})$。

(ii) 给定 $(\lambda, \theta)$ 后，各对 $(d_k, \{\eta_i\}^{i:k(i)=k})$、$k = 1, \ldots, K$ 相互独立。设 $s_i = \lambda_{k(i)} u_i^2/2$,

$$\pi(d_k \mid \lambda, \theta) \propto \pi_k(d_k) \prod_{i:k(i)=k} \mathcal{Z}_{H_{i,d_k}}(s_i), \tag{5}$$

并且，给定 $d_k$ 后，满足 $k(i) = k$ 的 $\eta_i$ 相互独立，其分布律为 $(H_{i,d_k})_{s_i}$。由于 $\mathrm{D}_k$ 有限，$d_k$ 可从 (5) 中精确抽取，随后从各自的倾斜分布中抽取权重。对于 Student-$t$族，$\mathrm{Z}_{H_d}(s) = (d/2)^{d/2}\Gamma(\frac{d+1}{2})\Gamma(\frac{d}{2})^{-1}(d/2+s)^{-(d+1)/2}$，且 $(H_d)_s =$服从 Gamma$(\frac{d+1}{2}, \frac{d}{2}+s)$；对于 slash族，$\mathrm{Z}_{H_\nu}(s) = \nu\,\Gamma(\nu+\frac{1}{2})P(\nu+\frac{1}{2}, s)\, s^{-(\nu+1/2)}$ 对于 $s > 0$，其中 $P$ 为正则化下不完全伽马函数，而 $(H_\nu)_s$ 服从截断在 (0, 1) 上的 Gamma$(\nu+\frac{1}{2}, s)$ 分布；对于 $s = 0$，$\mathrm{Z}_{H_\nu}(0) = \nu/(\nu+\frac{1}{2})$，且 $(H_\nu)_0 =$ 服从 Beta$(\nu+\frac{1}{2}, 1)$。

(iii) 给定 $(\lambda,\ d,\ \eta)$，$\theta$ 服从精度为 $P_{\lambda,\ \eta} = \lambda_e W^T W + \mathrm{diag}(\Sigma_\beta^{-1},\ \Lambda_{\lambda,\ \eta})$、均值为 $P_{\lambda,\ \eta}^{-1}\{\lambda_e W^T y + ((\Sigma_\beta^{-1}\mu_\beta)^T,\ 0^T)^T\}$ 的正常分布律，该分布律不依赖于 $d$。

设 $T_\lambda = \lambda_e X^T X + \Sigma_\beta^{-1}$、$M_\lambda = I_N - \lambda_e X T_\lambda^{-1} X^T$ 和 $Q_{\lambda,\ \eta} = \lambda_e Z^T M_\lambda Z + \Lambda_{\lambda,\ \eta}$。对 $P_{\lambda,\ \eta}$ 中的 $T_\lambda$ 取舒尔补，得到 $\mathrm{Var}(u \mid \lambda,\ \eta) = Q_{\lambda,\ \eta}^{-1}$。

### 2.3 马尔可夫链

算法1模拟马尔可夫链 $\Phi = \{(\lambda^{(n)}, d^{(n)}, \eta^{(n)}, \theta^{(n)})\}_{n\geq 0}$，其不变分布为后验分布。第一步不使用 $d$ 或先前的 $\lambda$，第二步仅使用 $(\lambda, \theta)$，第三步仅使用 $(\lambda, \eta)$。因此，$\{(\eta^{(n)}, \theta^{(n)})\}_{n\geq 0}$ 本身是一个马尔可夫链，我们称之为边际链。其状态空间为 $\mathsf{X} = \prod_{i=1}^{q} \mathsf{S}_i \times \mathbb{R}^{p+q}$，其中 $\mathsf{S}_i$ 在下述假设 (A0) 中定义。我们研究边际链，并通过引理 4 将结论推广到 $\Phi$。

## 3 主要结果

### 3.1 记号和基本假设

对于概率测度 $H$ 在 $\mathbb{R}_+$，$r \in \mathbb{R}$ 和 $c \in (0, 1]$, 记 $m_r(H) = \int \eta^r H(d\eta) \in (0,\ \infty]$ 并且，令$\mathrm{E}_s$表示在 $H_s$ 式（4）下的期望，

$$\kappa_H(c) = \limsup_{s\to\infty} \frac{\mathrm{E}_s[\eta^{-c}]}{s^c}, \qquad R_H(c) = \frac{\Gamma(c+\frac{1}{2})}{\Gamma(\frac{1}{2})}\,\kappa_H(c).$$

这些是de la Cruz 和 Humeniy (2026年)中的量。对于 Student-$t$ 族，其中有 $d$ 个自由度以及 $c < (d+1)/2$，$(H_d)_s =$伽马分布$(\frac{d+1}{2},\ \frac{d}{2}+s)$，$\mathrm{E}_s[\eta^{-c}] = \Gamma(\frac{d+1}{2}-c)\Gamma(\frac{d+1}{2})^{-1}(\frac{d}{2}+s)^{-c}$以及

$$R_{H_d}(c) = \frac{\Gamma(c+\frac{1}{2})\,\Gamma(\frac{d+1}{2}-c)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{d+1}{2})}.$$

对于参数为 $\nu$， $R_{H_\nu}(c)$ 的 slash族，其表达式与$(d+1)/2$ 相同，只需将其替换为 $\nu+\frac{1}{2}$，当 $c<\nu+\frac{1}{2}$ (de la Cruz 和 Humeniy，2026年，命题2(c))。二者均随尾部参数增大而递减。对于每个坐标 $i$ 令

$$R_i(c)=\max_{d\in\mathcal{D}_{k(i)}} R_{H_{i,d}}(c).$$

对于设计矩阵，令 $r=$rank$(Z)$，令 $Z_{-k}$ 表示去除分块 $Z_k$ 后的 $Z$， $r_{-k}=$ rank$(Z_{-k})$（当 $r_{-k}=0$ 为 $K=1$时），并令 $\Pi_Z$ 为 $\mathbb{R}^q$ 到 $\mathcal{N}(Z)=\{v:Zv=0\}$上的正交投影。定义

$$\ell_i=(\Pi_Z)_{ii},\qquad s_j(c)=\sum_{i=1}^{q}|(\Pi_Z)_{ij}|^{2c},\qquad S_k(c)=\max_{j:k(j)=k}s_j(c),\qquad \delta_k=q_k+r_{-k}-r.$$

因此 $0\le\ell_i\le 1$、 $\sum_i\ell_i=q-r$和 $s_j(1)=\ell_j$，从而 $S_k(1)=\ell_k^*:=\max_{i:k(i)=k}\ell_i$。数值 $\delta_k\ge 0$ 是 N $(Z)$ 在分块 $k$ 的坐标上的投影的维数（引理 5），因此当且仅当分块 $k$ 对应的 $\Pi_Z$ 的列为零时，$\delta_k=0$。我们还写作 $\tau_Z=$ tr$\{(Z^TZ)^+\}$，其中 $A^+$ 表示 Moore–Penrose逆。

我们始终作如下假设:

(A0) 对于每个 $i$，分布律 $H_{i,d}$ 和 $d\in\mathcal{D}^{k(i)}$ 相互绝对连续，并满足 $m_{1/2}(H_{i,d})<\infty$。我们将它们的公共支集记为 S$_i$。

对于任何网格上的 Student-$t$ 族，S$_i=\mathbb{R}_+$；对于 slash族，S$_i=(0,1]$。若 Harris遍历马尔可夫链的核为 $P$，不变分布为 $\pi$，且存在 $\rho<1$和函数 $C$满足 $\|P^n(x,\cdot)-\pi\|_{\mathrm{TV}}\le C(x)\rho^n$ 对所有 $x$和 $n$均成立，则该链是几何遍历的。

### 3.2 几何遍历性

**定理1**。 假设(A0) 和

(C1) $a_e>\frac{1}{2}(r-N+2)$；(C2) 对于每个满足 $\delta_k>0$的 $k$， $a_k>1-\frac{1}{2}(r-r_{-k})$。

假设存在 $c\in(0,1]$ 使得

(H1) $m_{c+1/2}(H_{i,d})<\infty$ 且$\kappa_{H_{i,d}}(c)<\infty$ 对每个 $i$ 及每个$\mathcal{D}^{k(i)}$均成立；

(H2) $R_j(c)s_j(c)<1$对于每个 $j=1,\ldots,q$。

则算法 $\Phi$ *1* 的 *Gibbs* 链是几何遍历的。

当混合分布仅取决于分量时，即对于 $k(i)=k$， $R_i(c)=R_k(c)$，(H2) 表示对于每个 $k$， $R_k(c)S_k(c)<1$，并且它仅限制具有 $\delta_k>0$ 的分量。当 $Z$ 具有满列秩时，(C2) 和 (H2) 均为空，关于混合分布的条件简化为 (H1)。对于正态随机效应（在 1 处的点质量），$\kappa_{H_i,d}(c)=0$。此时，条件 (C1) 与 Román 和 Hobert（2015）针对具有正规先验的分块吉布斯采样器所给出的相应条件一致，但不要求 $X$ 具有满列秩；当 $K>1$时，(C2) 不比他们的条件更强，并且可能严格更弱，因为 $\delta_k\le q-r$，而且 (C2) 仅施加于具有 $\delta_k>0$ 的分量。对于 $K=1$ 且尾部参数已知的情形，定理1 退化为 de la Cruz 和 Humeniy (2026) 的定理1。在 $c=1$下，(H2) 表示 $\ell_jR_j(1)<1$；对于网格最小值为 $d_{\min}>1$ 的 Student-$t$ 随机效应，这就是 $d_{\min}>1+\ell_j$。 $c$ 的较小取值会减弱 (H1) 中的矩条件。对于 Student-$t$ 和 slash 分布族， $R_H(c)\to 1$ 随 $c\downarrow 0$ 变化，这在 $R_H(1)>1$时有所帮助，即对于 $d<2$ 或 $\nu<1$；另一方面，当 $c$ 减小时，$s_j(c)$ 增大，并趋近于 $\Pi_Z$ 的第 $j$ 列中非零元素的数量。

### 3.3 尾部参数未知的族

命题 *1.* 假设分量 $k$ 具有尾部参数取自有限网格 $D_k$ 的 *Student-t* 或 *slash* 随机效应，即 $H_{i,\,d}=$ 伽马分布$(d/2, d/2)$，或 $H_{i,\,d}=$ 贝塔分布$(d, 1)$，其中 $k(i)=k$，并令 $d_{\min}=\min D_k$。则 *(A0)* 对这些坐标成立，$H_{i,\,d}$ 的所有正矩均有限，并且对于 $c<(d_{\min}+1)/2$（*Student-t*）或 $c<d_{\min}+\frac{1}{2}$（*slash*），$R_i(c)=R_{H_{d_{\min}}}(c)$。特别地：

*(a)* 当 $c=1$ 时，对于这些坐标，*(H1)–(H2)* 在 $d_{\min}>1+\ell_k^*$（*Student-t*）或 $d_{\min}>(1+\ell_k^*)/2$（*slash*）的情况下成立；*(b)* 当 $\delta_k=0$ 时，对于这些坐标，对于每个 $c\in(0,1]$ 具有 $c<(d_{\min}+1)/2$ 或$c<d_{\min}+\frac{1}{2}$），*(H1)–(H2)* 均成立，因此允许使用任意有限网格；*(c)* 如果对于分块 $k$，$\Pi_Z$ 的每一列均为零向量或坐标向量，使得 $S_k\le 1$（$Z$ 的零列），则对于这些坐标，当 $c$ 足够小时以及对于任意有限网格，*(H2)* 成立。

对于污染正态分布、具有 $\chi$ 的正态-*GIG* 分布、$\psi>0$，以及形状参数为 $\alpha>\frac{1}{2}$ 且参数已知的方差-伽马随机效应，$\kappa_H(c)=0$ 对每个 $c$ 均成立，因此 (H2) 对这些坐标成立，并且 (H1) 对每个 $c\in(0,1]$, 均成立，唯一例外是方差-伽马分布需要 $c<\alpha-\frac{1}{2}$（*de la Cruz* 和 *Humeniy*，*2026*年，命题 *2* 和推论 *1*）。

由于 (H1)–(H2) 必须对所有分量都以相同的 $c$ 成立，实际应用中会在网格$_j$ $R_j(c)^{s_j(c)}$ 上取一组 $c$值进行计算；第 5 节给出了一个例子。

### 3.4 设计重缩放

令 $C=\mathrm{diag}(c_1,\dots,c_q)$，其中 $c_i>0$；并令 $s_j(c;\,C)$ 按照 $s_j(c)$ 计算，其中将 $\Pi_Z$ 替换为 $\mathbb{R}^q$ 在 $C\mathcal{N}(Z)$ 上的正交投影。这是投影到 $ZC^{-1}$的零空间，即其中 $Z$ 的第 $i$列已除以 $c_i$的设计矩阵。

**推论 1.** 定理 *1*的结论仍然成立，只要将 (H2) 替换为：存在一个正对角矩阵 $C$ 使得 $R_j(c)^{s_j(c;C)}<1$ 对每个 $j$均成立。

该推论是等变性的一种结果。设计矩阵为$ZC^{-1}$、随机效应为 $Cu$ 且权重为 $C^{-2}\eta$ 的重缩放模型，是同一模型的一种等价参数化；其吉布斯采样器是算法 1 在一个固定双射下的映像，并且 (A0)、(C1)–(C2) 和(H1)中的量，以及 $R_j(c)$，在重缩放后均保持不变（第 4.8节）。然而，投影到零空间的投影会发生变化。

例1（成比例列）。假设 N (*Z*) 由向量 $w_1$、……、$w_m$ 张成，这些向量的支撑互不相交，且每个向量的支撑包含两个坐标：$w_j=$（……，$a_j$，……，$-b_j$，……），其中 $a_j$、$b_j\neq 0$。当唯一的线性依赖来自每次只有一次观测的受试者 $t\neq 0$时，这就是随机截距和斜率的情形：这类受试者的截距列和斜率列分别为 $e$和 $te$，而 N (*Z*) 在这两个坐标上的部分为 $(t,-1)$。对于每一对坐标，$\Pi_Z$是对角线上元素为 $\ell$、$1-\ell$，非对角线上元素为 $\pm\{\ell(1-\ell)\}^{1/2}$ 的 $2\times 2$ 矩阵，其中$\ell=a_j^2/(a_j^2+b_j^2)$；对于在时间 $t$仅观测一次的受试者的截距，$\ell=t^2/(1+t^2)$。令每对坐标中的第一个坐标取$c_i=|b_j/a_j|$，其他位置取 $c_i=1$，则每一对坐标均有 $\ell=\frac{1}{2}$，因此 $s_j(c;\,C)=2^{1-2c}$在这些成对坐标上成立，且 $s_j(1\,2;\,C)=1$。对于 Student-$t$族，$R_{H_d}(1\,2)=\Gamma(\frac{d}{2})/\{\Gamma(1\,2)\Gamma(\frac{d+1}{2})\}$，当且仅当 $d>1$时，该值小于 1；对于 slash族，$R_{H_\nu}(1\,2)<1$当且仅当 $\nu>\frac{1}{2}$。因此，对于 Student-$t$随机截距和斜率，取 $c=\frac{1}{2}$时，推论 1 表明，只要所有网格均满足 min $D_k>1$，无论……的数量如何，几何遍历性都成立

有单次观测的受试者，前提是其中没有受试者仅在时间零被观测。恰好在原点观测一次的受试者，其斜率列为零，因此 $e_i \in \mathcal{N}(Z)$ 且 $s_i(c;\ C) = 1$ 对每个 $C$均成立；对于该坐标，(H2) 仍可写作 $R_i(c) < 1$，在相同网格条件下，该式在 $c = \frac{1}{2}$ 处成立。

还可以进一步进行重缩放。若集合 $J$ 包含 $q - r$ 个坐标，且其补集中的$Z$ 列线性无关，则称该集合为 主元集；这样的集合总是存在，因为任意一组 $r$线性无关列的补集都是一个主元集。

**推论 2**。设 $J$ 为一个主元集。定理*1* 的结论在(H2)替换为以下条件时仍然成立：$R_j(c) < 1$ 对每个 $j \in J$均成立。

其思路是将 $J$ 中的坐标乘以一个较大的常数：重缩放后设计矩阵的零空间便集中于 $J$的坐标上，其投影趋近于在 $J$上取值为 1 的对角矩阵，而重缩放后的列表现得如同 $Z$ (第4.9节)。对于 $K = 1$以及已知尾部参数的情形，这是de la Cruz 和 Humeniy(2026年)的定理 2。结合命题1和推论2可知，$Z$ 的秩对常见分布族不构成任何限制。

**推论3**。 假设(A0) 和(C1)–(C2)成立，并且假设每个分量的随机效应服从 *Student-t* 分布或 *slash* 分布，且其尾部参数取自有限网格；或者服从污染正态分布、带有$\chi$、$\psi > 0$的正态-*GIG* 分布，或形状参数大于1 2 且参数已知的方差-伽马随机效应。那么，无论设计矩阵 $Z$如何，$\Phi$ 都是几何遍历的。

定理1 和推论1 在某个族满足 $R_H(c) \geq 1$ 对所有足够小的 $c$时仍然适用，并且说明了投影到零空间如何进入漂移项；对于上述各族，推论3使关于尾部的条件不再需要。

### 3.5 共享权重：多元尺度混合

在随机系数模型中，受试者的随机效应向量的各个坐标通常共享一个潜在权重，因此该向量服从多元正态尺度混合，例如多元 Student-$t$ (Pinheiro 等人，2001年)；在该模型中，受试者的误差项也共享这个权重，而这里的误差项服从正态分布。现在我们允许随机效应共享权重。将坐标$1, \ldots, q$ 划分为若干分块 $b \in \mathcal{B}$，每个分块都属于单个分量，记 $b(i)$ 为坐标 $i$ 所属的分块，$n_b$ 为分块 $b$ 的大小，$k(b)$ 为其所属分量，$u_b$ 为 $u$的对应子向量。模型（2）替换为

$$u_i \mid \lambda, \eta, d \overset{\text{ind}}{\sim} \mathrm{N}(0, (\lambda_{k(i)}\eta_{b(i)})^{-1}), \qquad \eta_b \mid d \overset{\text{ind}}{\sim} H_{b,d_{k(b)}}, \qquad b \in \mathcal{B}, \tag{6}$$

因此，$u_b$ 是 $\mathrm{N}^{n_b}(0,\ \lambda_{k\ b\ ()}^{-1} I)$ 分布律的尺度混合。模型（2）是块大小为 1 的情形。完全条件分布 (i) 和 (iii) 保持不变，其中以 $\eta_{b(i)}$ 替代 $\eta_i$。在 (ii) 中，当 $s_b = \lambda_{k(b)}\|u_b\|^2/2$ 时，权重 $\eta_b$ 服从倾斜分布 $(H_{b,\ d_{k(b)}})^{(n_b)s_b}$，其中，对于定义在 $\mathbb{R}_+$ 上的概率测度 $H$ 和 $n \geq 1$，

$$\mathcal{Z}_H^{(n)}(s) = \int_0^\infty \eta^{n/2} e^{-s\eta} H(d\eta), \qquad H_s^{(n)}(d\eta) = \frac{\eta^{n/2} e^{-s\eta} H(d\eta)}{\mathcal{Z}_H^{(n)}(s)},$$

并且在（5）中以 $\mathrm{Z}^{(n_b)}{}_{H_b,d}(s_b)$ 替换 $\mathrm{Z}_{H_i,d}(s_i)$。对于 $H_d =$ 伽马分布$(d/2, d/2)$，$(H_d)_s^{(n)} =$伽马分布$(\frac{d+n}{2}, \frac{d}{2} + s)$。令 $\kappa_H^{(n)}(c) = \limsup_{s\to\infty} s^{-c}\mathrm{E}_s^{(n)}[\eta^{-c}]$，其中 $\mathrm{E}_s^{(n)}$ 表示在 $H_s^{(n)}$ 下的期望，

$$R_H^{(n)}(c) = \frac{\Gamma(c + \frac{n}{2})}{\Gamma(\frac{n}{2})}\,\kappa_H^{(n)}(c), \qquad R_b(c) = \max_{d \in \mathcal{D}_{k(b)}} R_{H_{b,d}}^{(n_b)}(c),$$

因此 $R_H^{(1)} = R_H$。对于分块 $b$，$b'$令 $F_{bb'} = \sum_{i\in b,\ j\in b'} (\Pi_Z)_{ij}^2$ 表示 $\Pi_Z$ 中对应块的平方弗罗贝尼乌斯范数，并且

$$\sigma_{b'}(c) = \sum_{b\in\mathcal{B}} \frac{\vartheta_{n_b}(c)}{\vartheta_{n_{b'}}(c)} \Big(\frac{F_{bb'}}{n_{b'}}\Big)^c, \qquad \vartheta_n(c) = \frac{\Gamma(c+\frac{n}{2})}{\Gamma(\frac{n}{2})(\frac{n}{2})^c}.$$

当所有块的大小都相同时，$n$，$\sigma_{b'}(c) = \sum_b (F_{bb'}/n)^c$；对于 $n = 1$，则为 $s_{b'}(c)$，而一般情况下，$\sigma_{b'}(1) = n_{b'}^{-1} \sum_{j\in b'} \ell_j$。

定理 *2.* 考虑模型 *(1)*、*(6)*、*(3)*，并假设 *(C1)–(C2)* 以及 *(A0)* 的分块版本：对每个 $b$，分布律 $H_{b,d}$，$d \in \mathcal{D}^{k(b)}$ 相互绝对连续，且具有公共支集 $S_b$，并且 $m_{n_b/2}(H_{b,d}) < \infty$。边际链于是运行于 $\prod_b S_b \times \mathbb{R}^{p+q}$ 上。假设存在 $c \in (0,1]$，使得 $m_{c+n_b/2}(H_{b,d}) < \infty$ 且 $\kappa^{(n_b)}{}_{H_{b,d}}(c) < \infty$，对每个 $b$ 和 $d \in \mathcal{D}^{k(b)}$ 均成立，并且 $R_b(c)\,\sigma_b(c) < 1$ 对每个 $b \in \mathcal{B}$ 均成立。那么 *Gibbs* 链是几何遍历的。

该条件涉及 $\Gamma(c+\frac{n}{2})/\Gamma(\frac{n}{2})$，而不是 $\Gamma(c+\frac{1}{2})/\Gamma(\frac{1}{2})$，因为分块范数的分数阶矩是一个 $\chi^2_{n_b}$ 变量的分数阶矩，而倾斜分布会增加因子 $\eta^{n_b/2}$。最常见的情况是个体特异性设计。假设 $m$ 个受试者的随机效应向量为$v_b \in \mathbb{R}^{n_0}$，其设计矩阵为 $Z^{[b]}$ $(N \times n_0)$，其中非零行是受试者 $b$的设计矩阵行，因此$Zu = \sum_b Z^{[b]} v_b$，并且

$$v_b \mid \lambda, \eta, d \overset{\text{ind}}{\sim} \mathrm{N}_{n_0}\big(0, (\lambda_{g(b)}\eta_b \Psi_{g(b)})^{-1}\big), \qquad \eta_b \mid d \overset{\text{ind}}{\sim} H_{g(b),d_{g(b)}}, \tag{7}$$

其中，受试者被划分为若干组 $g$，每组具有一个尺度 $\lambda_g$、一个已知的正定形状矩阵 $\Psi_g$ 以及一个取自有限网格的尾部参数 $d_g$。对于 $H_d =$ 伽马分布$(d/2,\ d/2)$，$v_b$ 服从自由度为 $d_g$、尺度矩阵为 $(\lambda_g\Psi_g)^{-1}$的多元 Student-$t$ 分布。令 $\mathrm{nul}_b = n_0 -$ 秩$(Z^{[b]})$。

推论*4*。考虑模型*(1)*、*(7)*、*(3)*，将各组视为分量，并假设在$m_{1/2}$的位置使用$m_{n_0/2}(H_{g,d}) < \infty$，且满足*(C1)–(C2)*。假设对于每个组 $g$ 和 $d \in \mathcal{D}_g$，存在 $c \in (0,1]$ *with* $m_{c+n_0/2}(H_{g,d}) < \infty$, $\kappa^{(n_0)}{}_{H_g}$，以及$_d(c) < \infty$；并且对于每个满足 $\mathrm{nul}_b > 0$ 的受试者 $b$ 和每个 $d \in \mathcal{D}^{g(b)}$，有 $R^{(n_0)}{}_{H_{g(b)},d}(c)(\mathrm{nul}_b/n_0)^c < 1$。则 *Gibbs* 链是几何遍历的。特别地，对于具有任意有限网格尾部参数和任意个体特异性设计的多元 *Student-*$t$ 和 *slash*随机效应，该结论成立。

对于多元 Student-$t$随机效应，$\kappa^{(n_0)}{}_{H_d}(c) = \Gamma\big(\frac{d+n_0}{2} - c\big)/\Gamma\big(\frac{d+n_0}{2}\big)$ 对于 $c < \frac{d+n_0}{2}$，而对于 slash族，将 $\frac{d+n_0}{2}$替换为 $\nu + \frac{n_0}{2}$ 即可。无需重新缩放：对于个体特异性设计，$Z$ 的零空间是各个 $Z^{[b]}$零空间的直和，因此 $\Pi_Z$ 是分块对角的。形状矩阵 $\Psi_g$ 在（7）中是已知的；未知的非结构化形状情形将在第7节讨论。

### 3.6 后验矩

几何遍历性可为任何具有高于二阶后验矩的函数导出马尔可夫链中心极限定理，并在稍强的矩条件下得到渐近协方差的一致批均值估计量（Jones，2004年；Vats etal.，2019年；Banerjee and Vats，2024）。下一个结果为通常监测的量提供这些矩。

**命题 2。** 在模型 (1)–(3)中，以及第 (6) 节 (7) 和 *3.5*节中具有共享权重的模型中，对于所有混合分布，在后验分布下 $\mathrm{E}\|\beta\|^m < \infty$，$\mathrm{E}|\log \lambda_e|^m < \infty$ 以及 $\mathrm{E}|\log \lambda_k|^m < \infty$ 对每个 $m > 0$ 和每个 $k$均成立。

尾部参数取值于有限集，因此其所有矩均有限；而且，中心极限定理适用于有界的指示变量 $\mathbf{1}\{d_k = d\}$。在定理1、推论3、定理 2 或推论4的假设下，中心极限定理因而适用于 $g = (\beta^T, \log \sigma_e, \log \sigma_1, \ldots, \log \sigma_K)^T$，其中$\sigma_e = \lambda_e^{-1/2}$和 $\sigma_k = \lambda_k^{-1/2}$，并适用于尾部参数的后验概率。由于$g$ 具有各阶后验矩，对于每个 $t_b \in (1\,2, 1)$，批大小为$b_n = \lfloor n^{t_b} \rfloor$的多元批均值估计量均强相合。事实上，对于几何遍历链，若某函数的后验矩阶数大于四，则 Banerjee and Vats (2024) 的定理 2(b) 给出速率为 $n^{1/4} \log n$ 的强不变原理；Vats et al. (2019) 的强相合判据（Banerjee and Vats (2024) 的定理 3 对其作了重述）适用，因为$b_n^{-1} \log n\,(n^{1/4} \log n)^2 = O\{n^{1/2-t_b} (\log n)^3\}$ 趋于零，即对于 $t_b > \frac{1}{2}$。

# 4 证明

向量的范数均为欧几里得范数，且 $\|A\|_2$ 为谱范数。对于对称矩阵，$A \preceq B$ 表示 $B - A$ 为半正定。

## 4.1 单分量情形的结果

以下三个结果是 de la Cruz 和 Humeniy (2026年) 的引理2、5和4；为使本文自成一体，相关证明见附录 B。第一个结果将设计矩阵的行空间与其零空间分离。

引理 *1*（*de la Cruz* 和 *Humeniy*，*2026*年，引理 *2*）。设 $A$ 为一个 $m \times n$ 矩阵，$\upsilon > 0$，$\Upsilon = \mathrm{diag}(\upsilon_1, \ldots, \upsilon_n)$，其中 $\upsilon_i > 0$，且 $\Omega = \upsilon A^T A + \Upsilon$。设 $\Pi_A$ 为到 $\tilde{N}(A)$ 上的正交投影，$P_A = I - \Pi_A$，且 $\ell_i = (\Pi_A)_{ii}$。则

(a) $P_A Q^{-1} P_A \preceq \upsilon^{-1} (A^T A)^+$ 且 $\Pi_A Q^{-1} \Pi_A \preceq \Pi_A \Upsilon^{-1} \Pi_A$ 对于每个 $Q \succeq \Omega$；

(b) $\mathrm{tr}(\Omega^{-1}) \le \upsilon^{-1} \mathrm{tr}\{(A^T A)^+\} + \sum_{i=1}^n \tilde{\ell}_i \upsilon_i^{-1}$；(c) $\mathrm{tr}(A\Omega^{-1} A^T) \le \upsilon^{-1}$ 秩$(A)$。

第二个涉及倾斜族（4）。

**引理 2** (de la Cruz 和 Humeniy，2026年，引理 5)。设 $H$ 是定义在 $\mathbb{R}_+$ 上的概率测度，满足 $m_{1/2}(H) < \infty$，且 $c \in (0, 1]$。

(a) $\sup_{s \ge 0} \mathrm{E}_s[\eta^c] = m_{c+1/2}(H)/m_{1/2}(H)$。

(b) 如果 $\kappa_H(c) < \infty$，那么对于每个 $\kappa' > \kappa_H(c)$，都存在 $C < \infty$，使得 $\mathrm{E}_s[\eta^{-c}] \le \kappa' s^c + C$ 对于所有 $s \ge 0$.

第三个界对权重取均匀上界，给出了 $\theta$ 的条件均值。其证明将 $\mathrm{E}(\theta \mid \lambda, \eta)$ 写成加权最小二乘解，并使用 Stewart (1989年) 和 O'Leary(1990年) 关于缩放投影的界。它只使用了 $u$ 的先验精度矩阵是正对角矩阵这一点，因此对 $\Lambda_{\lambda,\,\eta}$ 同样直接适用。

引理 *3*（*de la Cruz* 和 *Humeniy*，*2026*年，引理 *4*）。存在有限常数 $K_1$ 和 $K_2$，它们仅依赖于 $X$、$Z$、$y$、$\mu_\beta$ 和 $\Sigma_\beta$，使得对于所有 $\lambda \in \mathbb{R}_+^{K+1}$ 和 $\eta \in \mathbb{R}_+^q$，均有 $\|E(\theta \mid \lambda, \eta)\| \le K_1$ 以及 $\|y - WE(\theta \mid \lambda, \eta)\| \le K_2$。

最后，我们使用以下基本事实（de la Cruz 和 Humeniy，2026年，引理 6），这些事实也在附录B中得到证明。对于 $p > 0$ 和 $\epsilon > 0$，存在 $C_{\epsilon,\,p}$，使得 $(x + y)^p \le (1 + \epsilon) x^p + C_{\epsilon,\,p} y^p$对 $x$、$y \ge 0$成立。若 $Y \sim \mathrm{N}(m, \sigma^2)$ 且 $c \in (0, 1]$，则 $\mathrm{E}|Y|^{2c} \le (1 + \epsilon)\mu_c \sigma^{2c} + C_{\epsilon,2c} |m|^{2c}$，其中$\mu_c = \mathrm{E}|\xi|^{2c} = 2^c \Gamma(c + \frac{1}{2})/\Gamma(1\,2)$ 对 $\xi \sim \mathrm{N}(0, 1)$ 成立。我们还使用 $(\sum_j w_j)^c \le \sum_j w_j^c$对 $w_j \ge 0$成立，以及詹森不等式 $\mathrm{E}[X^c] \le (\mathrm{E}X)^c$ 对 $X \ge 0$成立。

## 4.2 拓扑性质

令 $P_m$ 为边际链的核， $P_\Phi$ 为 $\Phi$的， $\pi$ 为后验分布，且 $\pi_m$ 其 $(\eta,\ \theta)$对于每个 $i$ 固定 $d_i^\circ \in \mathcal{D}^{k(i)}$，令 $H_i = H_{i,\ d_i^\circ}$ 并根据 (A0)，令$h_{i,\ d} = dH_{i,\ d}/dH_i$，其为正_ _$H_i$-几乎处处。令 $\varphi = \bigotimes_i H_i \otimes \mathrm{Leb}^{p+q}$。

**引理 4**。 在(A0)条件下，边际链和 $\Phi$ 是 *Harris*遍历的，并且X 的每个紧子集对于边际链都是小集。如果边际链是几何遍历的，则 $\Phi$也是如此。

证明。算法1的每一步都从其完全条件分布中抽样一个分块，因此 $\pi$ 对于 $P_\Phi$ 是不变的，而$\pi_m$ 对于 $P_m$是不变的。从（$\eta', \theta'$）出发，边际链的一步先抽样 $\lambda \sim \pi$（$\lambda \mid \eta', \theta'$），然后从给定（$\lambda, \theta'$）的联合完全条件分布中抽样（$d, \eta$），再抽样 $\theta \sim \pi$（$\theta \mid \lambda, \eta$）。对 $d$,¯求和后，给定（$\lambda, \theta'$）时 $\eta$ 的分布律相对于 $\bigotimes_i H_i$具有密度

$$g(\eta \mid \lambda, \theta') = \prod_{k=1}^{K} \sum_{d \in \mathcal{D}_k} \pi(d_k = d \mid \lambda, \theta') \prod_{i:k(i)=k} \frac{h_{i,d}(\eta_i)\, \eta_i^{1/2} e^{-s_i' \eta_i}}{\mathcal{Z}_{H_{i,d}}(s_i')}, \qquad s_i' = \tfrac{1}{2} \lambda_{k(i)} u_i'^2,$$

因此 $P_m$ 具有相对于 $\varphi$的密度 $k$（$\eta, \theta \mid \eta', \theta'$） $= \int \pi$（$\theta \mid \lambda, \eta$） $g$（$\eta \mid \lambda, \theta'$） $\pi$（$\lambda \mid \eta', \theta'$） $d\lambda$，该密度在 $\varphi$-几乎处处为正，因为每个因子都是正的。因此，只要 $\varphi$（$A$） $> 0$，对于每个 $x \in \mathsf{X}$，$P_m$（$x, A$） $> 0$，且该链是 $\varphi$-不可约的。它是非周期的：如果 $D_1, \ldots, D_\varpi$与 $\varpi \geq 2$构成一个周期循环，则 $\varphi$（$D_1$） $> 0$，因为对于 $x \in D_\varpi$， $P_m$（$x, D_1$） $= 1$，且 $P_m$（$x, \cdot$）相对于 $\varphi$具有密度；而对于 $x \in D_1$， $P_m$（$x, D_1$） $= 0$，矛盾。相对于$\mathrm{Leb}_{K+1} \otimes$（计数测度）的后验密度 $\otimes \bigotimes_i H_{i,d_{k(i)}} \otimes \mathrm{Leb}^{p+q}$为正，并且根据（A0）， $\pi_m$与 $\varphi$等价。因此，对于每个 $x$， $P_m$（$x, \cdot$） $\ll \pi_m$，而Tierney（1994）的推论1给出Harris常返性；结合不变概率测度 $\pi_m$，该链是Harris遍历的。同样的论证适用于 $\Phi$，其核相对于$\mathrm{Leb}_{K+1} \otimes$（计数测度） $\otimes \varphi$具有正密度。

对于 Feller性质，我们证明 $k$ 关于 $(\eta', \theta')$ 连续，且在局部上一致地被某个可积函数所控制。伽马密度 $\pi(\lambda \mid \eta', \theta')$ 关于 $(\eta', \theta')$ 连续变化，其率参数下界分别为 $b_e$ 和 $b_k$，并且在紧集上有上界。权重 $\pi(d_k = d \mid \lambda, \theta') \in [0, 1]$ 关于 $(\lambda, \theta')$ 连续，因为每个 $\mathsf{Z}_{H_i,d}$ 都连续且为正。固定目标点 $(\eta, \theta)$，其中 $h_{i,d}(\eta_i) < \infty$ 对每个 $i$ 和 $d$ 都成立；这对 $\varphi$-几乎处处的目标点成立。于是，仅依赖于 $(\eta, \theta)$ 的因子为常数。对于倾斜因子，令 $\delta > 0$。由于 $\eta_i \in \mathsf{S}_i$，即 $H_{i,d}$ 的支撑，我们有 $\mathsf{Z}_{H_i,d}(s) \geq e^{-s(\eta_i+\delta)} \int_{(0,\eta_i+\delta)} t^{1/2} H_{i,d}(dt)$，其积分为正，因此每个因子都至多为 $C_\delta(\eta_i) h_{i,d}(\eta_i) \eta_i^{1/2} e^{s_i'\delta}$，其中 $C_\delta(\eta_i)$ 是对有限集 $\mathsf{D}^{k(i)}$ 取最大值，且不依赖于 $(\lambda, \eta', \theta')$。在 $(\eta', \theta')$ 的一个紧邻域上，若对所有 $i$ 都有 $|u_i'| \leq \bar{u}$，则取 $\delta$，使得 $q\delta\bar{u}^2 < \min_k b_k$；于是对 $i$ 的乘积至多为一个常数乘以 $\exp(\frac{1}{2}\sum_{k=1}^K b_k \lambda_k)$，该常数依赖于固定的 $\eta$。记 $\lambda_0 = \lambda_e$, $b_0 = b_e$, $a_0^\star = a_e + \frac{N}{2}$ 和 $a_k^\star = a_k + \frac{q_k}{2}$。 $\lambda_k$ 的完全条件分布为 $\mathrm{Gamma}(a_k^\star, \beta_k')$，其率参数 $\beta_k' \geq b_k$ 在该邻域上被某个 $\bar{\beta}_k$ 所界，因此其密度至多为 $\bar{\beta}_k^{a_k^\star} \Gamma(a_k^\star)^{-1} \lambda_k^{a_k^\star - 1} e^{-b_k \lambda_k}$；并且，对于某些依赖于固定目标点的 $C$ 和 $m$，有 $\pi(\theta \mid \lambda, \eta) \leq (2\pi)^{-(p+q)/2} \det(P_{\lambda,\eta})^{1/2} \leq C(1 + \sum_{k=0}^K \lambda_k)^m$，因为 $P_{\lambda,\eta}$ 关于 $\lambda$ 是仿射的。因此， $\lambda$ 中的被积函数在该邻域上一致地被以下函数控制

$$C'\Big(1 + \sum_{k=0}^{K} \lambda_k\Big)^m \prod_{k=0}^{K} \lambda_k^{a_k^\star - 1} \exp\Big(-\frac{1}{2}\sum_{k=0}^{K} b_k \lambda_k\Big),$$

它在 $\mathbb{R}_+^{K+1}$ 上可积，因为每个 $a_k^\star$ 都为正，且尾部呈指数衰减。由 $\lambda$ 中的控制收敛定理， $k(\eta,\ \theta \mid \eta',\ \theta')$ 对于 $\varphi$-几乎处处的 $(\eta,\ \theta)$，关于 $(\eta',\ \theta')$ 连续，并且根据

谢费引理 $(\eta', \theta') \mapsto P_m((\eta', \theta'), \cdot)$ 在全变差意义下连续，因此该链是弱费勒链。由于 $\mathsf{X}$ 是 $\mathbb{R}_+^q \times \mathbb{R}^{p+q}$的闭子集，因此它是局部紧且可分的，并且 $\operatorname{supp}(\varphi) = \mathsf{X}$。Meyn 和 Tweedie（1993年）的定理 6.2.9 和 6.2.5 表明，该链是一个 T 链，且每个紧集都是小集。

最后，$\Phi^{(n+1)}$ 由 $(\eta^{(n)}, \theta^{(n)})$ 通过一个固定的核 $R$ 得到，且 $\pi_m R = \pi$。由于应用马尔可夫核不会增大全变差距离，对于 $x = (\lambda, d, \eta, \theta)$ 和 $n \geq 0$，有 $\|P_\Phi^{n+1}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \|P_m^n((\eta, \theta), \cdot) - \pi_m\|_{\mathrm{TV}}$。如果右侧至多为 $C_m(\eta, \theta)\rho^n$，则对所有 $n \geq 0$，有 $\|P_\Phi^n(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \max\{1, C_m(\eta, \theta)/\rho\}\rho^n$。 □

根据引理 4 以及 Meyn 和 Tweedie (1993年）中的引理 15.2.8 和定理 15.0.1，定理 1成立，只要我们找到 $v : \mathsf{X} \to [1, \infty)$ 具有紧次水平集，且对某些 $\rho < 1$ 和 $L < \infty$，

$$\mathrm{E}\big[v(\eta, \theta) \mid \eta', \theta'\big] \leq \rho\, v(\eta', \theta') + L \qquad \text{for all } (\eta', \theta') \in \mathsf{X}, \tag{8}$$

其中，期望是相对于 $P_m((\eta', \theta'), \cdot)$ 而取的。

### 4.3 分块杠杆值

**引理 5。** 设 $\Lambda$ 为正对角 $q \times q$ 矩阵，设 $S = \Lambda N(Z)$，并设 $\Pi_S$ 为到 $S$ 上的正交投影。对于每个 $k$、$\sum_{i:\ k(i)=k}(\Pi_S)_{ii} \leq \delta_k$，且 $\delta_k$ 是 $N(Z)$ 在分块 $k$的坐标上的投影的维数。

证明。设 $P_k$ 为到分块 $k$的坐标上的正交投影。则 $\sum_{i:k(i)=k}(\Pi_S)_{ii} = \mathrm{tr}(P_k \Pi_S P_k)$。矩阵 $P_k \Pi_S P_k$ 是半正定的，其特征值属于 $[0, 1]$,，因此其迹至多等于其秩，而其秩至多为 $\mathrm{rank}(P_k \Pi_S) = \dim P_k \mathsf{S}$。由于 $\Lambda$ 是对角且可逆的，它与 $P_k$ 可交换，并且 $\dim P_k \mathsf{S} = \dim P_k \Lambda \mathsf{N}(Z) = \dim \Lambda P_k \mathsf{N}(Z) = \dim P_k \mathsf{N}(Z)$。最后，将映射 $v \mapsto P_k v$ 限制在 $\mathsf{N}(Z)$ 上所得的核由在分块 $k$ 上为零的 $v \in \mathsf{N}(Z)$ 构成，这些元素对应于 $\mathsf{N}(Z_{-k})$ 中的元素；因此 $\dim P_k \mathsf{N}(Z) = \dim \mathsf{N}(Z) - \dim \mathsf{N}(Z_{-k}) = (q - r) - (q - q_k - r_{-k}) = \delta_k$。 □

### 4.4 条件矩

令 $\varsigma_{\max}$ 为 $X\Sigma_\beta^{1/2}$、$c_\lambda^{-1} = \lambda_e^{-1} + \varsigma_{\max}^2$ 和 $\Lambda_\lambda = \mathrm{diag}(\lambda_{k(1)}, \ldots\ldots, \lambda_{k(q)})$ 的最大奇异值。与de la Cruz 和 Humeniy (2026年，第4节.2) 中一样，$(1 + \lambda_e \varsigma_{\max}^2)^{-1} I \preceq M_\lambda \preceq I$ 和 $XT_\lambda^{-1}X^T \preceq X\Sigma_\beta X^T$，因此 $\lambda_e Z^T M_\lambda Z \succeq c_\lambda Z^T Z$ 和 $Q_{\lambda,\ \eta} \succeq c_\lambda Z^T Z + \Lambda_{\lambda,\ \eta}$。

**引理 6.** 对于所有 $\lambda \in \mathbb{R}_+^{K+1}$ 和 $\eta \in \mathbb{R}_+^q$，

(a) $\mathrm{tr}(D_\eta^{1/2} Q_{\lambda,\ \eta}^{-1} D_\eta^{1/2}) \leq c_\lambda^{-1} \tau_Z \max_i \eta_i + \sum_{k=1}^K \delta_k \lambda_k^{-1}$;

(b) $\mathrm{tr}\{\mathrm{Var}(W\theta \mid \lambda, \eta)\} \leq c_\lambda^{-1} r + \mathrm{tr}(X\Sigma_\beta X^T)$.

证明。(a) 我们有 $D_\eta^{1/2} Q_{\lambda,\eta}^{-1} D_\eta^{1/2} = (D_\eta^{-1/2} Q_{\lambda,\eta} D_\eta^{-1/2})^{-1} \preceq (c_\lambda D_\eta^{-1/2} Z^T Z D_\eta^{-1/2} + \Lambda_\lambda)^{-1}$。应用引理 1(b)，其中 $A = ZD_\eta^{-1/2}$，$v = c_\lambda$ 和 $\Upsilon = \Lambda_\lambda$。$A$ 的零空间为 $D_\eta^{1/2}\mathsf{N}(Z)$，因此根据引理 5，取 $\Lambda = D_\eta^{1/2}$，$A$ 的杠杆值 $\tilde{\ell}_i$ 满足 $\sum_{i:k(i)=k} \tilde{\ell}_i \leq \delta_k$，且 $\sum_i \tilde{\ell}_i \lambda_{k\,i}^{-1}() \leq \sum_k \delta_k \lambda_k^{-1}$。此外，$\mathrm{tr}\{(A^T A)^+\} = \mathrm{tr}\{(AA^T)^+\} = \mathrm{tr}\{(ZD_\eta^{-1} Z^T)^+\}$。矩阵 $ZD_\eta^{-1}Z^T$和 $ZZ^T$具有相同的值域，且 $ZD_\eta^{-1}Z^T \succeq (\max_i \eta_i)^{-1} ZZ^T$；将二者限制在共同值域上，此时它们均为正定矩阵，取逆可得 $(ZD_\eta^{-1}Z^T)^+ \preceq \max_i \eta_i (ZZ^T)^+$。由于 $\mathrm{tr}\{(ZZ^T)^+\} = \tau_Z$，(a) 得证。

(b) 给定 $(\lambda,\eta,u)$，$\beta$服从精度为 $T_\lambda$的正态分布，且 $\mathrm{E}(X\beta+Zu\mid\lambda,\eta,u)$ 等于一个常数加上 $M_\lambda Zu$。根据全方差公式，$\mathrm{Var}(W\theta\mid\lambda,\eta)=XT_\lambda^{-1}X^T+M_\lambda ZQ_{\lambda,\eta}^{-1}Z^TM_\lambda$。第一项的迹至多为 $\mathrm{tr}(X\Sigma_\beta X^T)$。由于 $0\preceq M_\lambda^2\preceq I$，第二项的迹至多为 $\mathrm{tr}(ZQ_{\lambda,\eta}^{-1}Z^T)\le\mathrm{tr}\{Z(c_\lambda Z^TZ+\Lambda_{\lambda,\eta})^{-1}Z^T\}\le c_\lambda^{-1}r$，由引理 1(c) 可得。 □

## 4.5 定理 1 的证明

固定满足 (H1)–(H2) 的 $c$，并令 $\mathrm{I}=\{(i,d):1\le i\le q,\ d\in\mathrm{D}^{k(i)}\}$，这是一个有限集。由于 $m_{1/2}\le 1+m_{c+1/2}$，(A0) 和引理 2 适用于每个 $H_{i,d}$。令 $M_c=\max_{(i,d)\in\mathcal{I}}m_{c+1/2}(H_{i,d})/m_{1/2}(H_{i,d})<\infty$，且 $\kappa_i(c)=\max_{d\in\mathcal{D}_{k(i)}}\kappa_{H_{i,d}}(c)$。令 $\epsilon>0$ 和 $\kappa_i'=\kappa_i(c)+\epsilon$。根据 (H2)，我们可以选择足够小的 $\epsilon$，使得

$$\bar\gamma:=(1+\epsilon)^2\,\frac{\Gamma(c+\frac12)}{\Gamma(\frac12)}\,\max_j\kappa_j' s_j(c)<1.$$

由引理 2，对每个 $H_{i,d}$ 应用引理，并取 $\mathcal{I}$ 上各常数的最大值，可得存在 $C_0<\infty$，使得对所有 $(i,d)\in\mathcal{I}$ 和 $s\ge0$，$\mathrm{E}_{(H_{i,d})_s}[\eta^c]\le M_c$ 且 $\mathrm{E}_{(H_{i,d})_s}[\eta^{-c}]\le\kappa_i' s^c+C_0$。

固定 $(\eta',\theta')$，并将 $\mathrm{E}'$ 记为从 $(\eta',\theta')$ 出发、经过边际链一步后的期望。给定 $(\lambda,\theta')$，$\eta_j$ 的分布律是关于 $d_{k(j)}$ 的混合分布，其组成分布为 $(H_{j,d})_{s_j'}$，$s_j'=\lambda_{k(j)}u_j'^2/2$，且不涉及 $\lambda_e$；因此

$$\mathrm{E}[\eta_j^c\mid\lambda,\theta']\le M_c,\qquad \lambda_{k(j)}^{-c}\mathrm{E}[\eta_j^{-c}\mid\lambda,\theta']\le\kappa_j'2^{-c}|u_j'|^{2c}+C_0\lambda_{k(j)}^{-c}. \tag{9}$$

由伽马全条件分布和詹森不等式，

$$\mathrm{E}'[\lambda_e^{-c}]\le\Big(\frac{2b_e+\|y-W\theta'\|^2}{2a_e+N-2}\Big)^c,\qquad \mathrm{E}'[\lambda_k^{-c}]\le\Big(\frac{2b_k+\|D_{\eta',k}^{1/2}u_k'\|^2}{2a_k+q_k-2}\Big)^c,\qquad \mathrm{E}'[\lambda_k^c]\le\zeta_k^c, \tag{10}$$

其中 $\zeta_k=(2a_k+q_k)/(2b_k)$，$u_k$ 和 $D_{\eta,k}$ 分别是 $u$ 和 $D_\eta$ 中对应于分块 $k$的子向量和子矩阵。第一个界成立是因为由 (C1) 可知 $2a_e+N-2>r\ge0$。第二个界仅用于满足 $\delta_k>0$的 $k$，此时由 (C2) 和 $r-r_{-k}\le q_k$ 可得 $2a_k+q_k-2>q_k-(r-r_{-k})\ge0$。此外，$c_\lambda^{-c}\le\lambda_e^{-c}+\varsigma_{\max}^{2c}$。令

$$v(\eta,\theta)=\alpha_1\sum_{i=1}^q|u_i|^{2c}+\alpha_2\|D_\eta^{1/2}u\|^{2c}+\alpha_3\|y-W\theta\|^{2c}+\sum_{i=1}^q(\eta_i^c+\eta_i^{-c})+\|\beta\|^{2c},$$

其中 $\alpha_1$，$\alpha_3>0$ 待定；若某个 $\delta_k>0$，则 $\alpha_2>0$ 待定；否则为 $\alpha_2=0$。将不带系数的五项记作$V_1$、……、$V_5$，将它们在 $(\eta',\theta')$ 处的值记作 $V_{\mathit{1}}$。令$\mathcal{K}_+=\{k:\delta_k>0\}$。于是 $v\ge2q\ge1$，并且下文的常数 $C_1$，$C_2$，……以及 $L_1$，$L_2$，……均不依赖于 $(\eta',\theta')$。

次水平集是紧集。设 $\bar v\ge1$。若 $v(\eta,\theta)\le\bar v$，则每个 $\eta_i$ 都属于 $[\bar v^{-1/c},\bar v^{1/c}]$，$|u_i|^{2c}\le\bar v/\alpha_1$，对任意 $i$ 和 $\|\beta\|^{2c}\le\bar v$ 均成立。集合 $\{v\le\bar v\}$ 在 $\mathrm{X}$ 中是闭的，因为 $v$ 是连续的；并且它包含于 $\prod_i([\bar v^{-1/c},\bar v^{1/c}]\cap\mathrm{S}_i)\times\bar B$，其中 $\bar B$ 是一个闭球。该集合在 $\mathbb{R}^{2q+p}$ 中闭且有界，因为每个 $\mathrm{S}_i$ 在 $\mathbb{R}_+$ 中都是闭的。因此，$\{v\le\bar v\}$ 是紧集。

项 $V_1$。记 $\Pi=\Pi_Z$ 和 $P=I-\Pi$。给定 $(\lambda,\eta)$，由引理 3 可知 $u_i\sim\mathrm{N}(m_i,\sigma_i^2)$，其中 $|m_i|\le K_1$，且 $\sigma_i=\|Q_{\lambda,\eta}^{-1/2}e_i\|\le\|Q_{\lambda,\eta}^{-1/2}\Pi e_i\|+\|Q_{\lambda,\eta}^{-1/2}Pe_i\|$。引理 1(a) 中，取 $A=Z$，$v=c_\lambda$，$\Upsilon=\Lambda_{\lambda,\eta}$ 和 $Q=Q_{\lambda,\eta}$，可得 $e_i^T\Pi Q_{\lambda,\eta}^{-1}\Pi e_i\le\sum_j\Pi_{ij}^2(\lambda_{k(j)}\eta_j)^{-1}$ 以及 $e_i^TPQ_{\lambda,\eta}^{-1}Pe_i\le c_\lambda^{-1}\{(Z^TZ)^+\}_{ii}$。因此

$$\sigma_i^{2c}\le(1+\epsilon)\sum_j|\Pi_{ij}|^{2c}\lambda_{k(j)}^{-c}\eta_j^{-c}+C_1c_\lambda^{-c},$$

$$\mathrm{E}(V_1\mid\lambda,\eta)\le(1+\epsilon)^2\mu_c\sum_j s_j(c)\lambda_{k(j)}^{-c}\eta_j^{-c}+C_2c_\lambda^{-c}+C_3.$$

一列 $j$的 $\Pi$仅当 $\delta_{k(j)} > 0$时才非零。由（9）和 $\mu_c 2^{-c} = \Gamma\left(c + \frac{1}{2}\right) / \Gamma\left(\frac{1}{2}\right)$，

$$\mathrm{E}(V_1 \mid \lambda, \theta') \le \bar{\gamma} V_1' + C_4 \sum_{k:\delta_k>0} \lambda_k^{-c} + C_2 c_\lambda^{-c} + C_3.$$

由（10）、$(x+y)^c \le x^c + y^c$ 以及 $\|D_{\eta'}^{1/2}\, {}_k u_k'\| \le \|D_{\eta'}^{1/2} u'\|$,

$$\mathrm{E}' V_1 \le \bar{\gamma} V_1' + A_{12} V_2' + A_{13} V_3' + L_1$$

其中 $A_{12}$ 和 $A_{13}$为有限常数（若 $\mathcal{K}_+ = \emptyset$则 $A_{12}$ 为零）。

项 $V_2$ （仅当某个 $\delta_k > 0$时）。由引理 6(a) 和 3，$\max_i \eta_i \le \sum_i \eta_i$ 以及 $\delta_k = 0$ 对于$k \notin \mathcal{K}_+$,

$$\mathrm{E}(\|D_\eta^{1/2} u\|^2 \mid \lambda, \eta) \le (c_\lambda^{-1} \tau_Z + K_1^2) \sum_i \eta_i + \sum_{k \in \mathcal{K}_+} \delta_k \lambda_k^{-1},$$

由詹森不等式和 $(x+y)^c \le x^c + y^c$,

$$\mathrm{E}(V_2 \mid \lambda, \eta) \le (c_\lambda^{-c} \tau_Z^c + K_1^{2c}) \sum_i \eta_i^c + \Big( \sum_{k \in \mathcal{K}_+} \delta_k \lambda_k^{-1} \Big)^c.$$

第一项的 E′-期望至多为 $qM_c(\tau_Z^c \mathrm{E}'[c_\lambda^{-c}] + K_1^{2c}) \le A_{23} V_{\prime 3} + C_5$，由（9）和（10）可得。对于第二项，关于在 $\lambda$给定 $(\eta',\ \theta')$条件下 $\lambda$的分布律应用詹森不等式，并结合 $\mathrm{E}'[\lambda_k^{-1}]$的公式（当 $k \in \mathcal{K}_+$ 时该公式有限，依据 (C2)），即可得到

$$\mathrm{E}'\Big( \sum_{k \in \mathcal{K}_+} \delta_k \lambda_k^{-1} \Big)^c \le \Big( \sum_{k \in \mathcal{K}_+} \delta_k \frac{2b_k + \|D_{\eta',k}^{1/2} u_k'\|^2}{2a_k + q_k - 2} \Big)^c \le C_6 + A_{22} V_2',$$

with

$$A_{22} = \Big( \max_{k \in \mathcal{K}_+} \frac{\delta_k}{2a_k + q_k - 2} \Big)^c,$$

因为$\sum_{k \in \mathcal{K}_+} \|D_{\eta'}^{1/2}\, {}_k u_k'\|^2 \le \|D_{\eta'}^{1/2} u'\|^2$。对于 $k \in \mathcal{K}_+$，条件(C2) 等价于 $\delta_k < 2a_k + q_k - 2$，因此 $A_{22} < 1$，并且

$$\mathrm{E}' V_2 \le A_{22} V_2' + A_{23} V_3' + L_2.$$

项 $V_3$。由引理6(b) 和3可得，$\mathrm{E}(\|y - W\theta\|^2 \mid \lambda,\ \eta) \le c_\lambda^{-1} r + \mathrm{tr}(X \Sigma_\beta X^T) + K_2^2$。由詹森不等式，$(x+y)^c \le x^c + y^c$ 以及 (10),

$$\mathrm{E}' V_3 \le r^c \mathrm{E}'[\lambda_e^{-c}] + C_7 \le A_{33} V_3' + L_3, \qquad A_{33} = \Big( \frac{r}{2a_e + N - 2} \Big)^c < 1$$

由 (C1)。

项 $V_4$。由 (9) 和 (10),

$$\mathrm{E}' V_4 \le qM_c + \sum_j \left( \kappa_j' 2^{-c} \mathrm{E}'[\lambda_{k(j)}^c]\, |u_j'|^{2c} + C_0 \right) \le \bar{\kappa} 2^{-c} \bar{\zeta}^c V_1' + L_4,$$

其中 $\bar{\kappa} = \max_j \kappa_j'$ 和 $\zeta = \max_k \zeta_k$。

项 $V_5$。</style>给定 $(,\ \eta)$，根据引理 3，$\beta$ 服从均值范数不超过 $K_1$ 的正态分布，其协方差矩阵在半正定序下不超过 $\Sigma_\beta$，这是因为 $P_{\lambda,\ \eta} \succeq \mathrm{diag}(\Sigma_\beta^{-1},\ \Lambda_{\lambda,\ \eta})$。将正态 $Y$的 $\mathrm{E}|Y|^{2c}$ 上界应用于每个坐标，并结合 $(\sum_j w_j)^c \le \sum_j w_j^c$，可得 $\mathrm{E}(V_5 \mid \lambda,\ \eta) \le L_5$，其中 $L_5$为某个常数，因此 $\mathrm{E}' V_5 \le L_5$。无需对 $X$ 的秩作任何假设。

结论。 合并各项，

$$\mathrm{E}'v \le \Big(\bar\gamma + \frac{\bar\kappa 2^{-c}\bar\zeta^{c}}{\alpha_1}\Big)\alpha_1 V_1' + \Big(A_{22} + \frac{\alpha_1 A_{12}}{\alpha_2}\Big)\alpha_2 V_2' + \Big(A_{33} + \frac{\alpha_1 A_{13} + \alpha_2 A_{23}}{\alpha_3}\Big)\alpha_3 V_3' + L,$$

其中，若 $\mathrm{K}_+ = \emptyset$，则第二项不存在。若 $\mathrm{K}_+ \neq \emptyset$，则先选取 $\alpha_1 > \bar\kappa 2^{-c}\zeta^c/(1-\bar\gamma)$，再选取 $\alpha_2 > \alpha_1 A_{12}/(1-A_{22})$，然后选取 $\alpha_3 > (\alpha_1 A_{13} + \alpha_2 A_{23})/(1-A_{33})$。若 $\mathrm{K}_+ = \emptyset$，则 $\alpha_2 = 0$, $A_{12} = 0$, 项 $V_2$ 和常数 $A_{22}$、$A_{23}$ 均不出现，此时选择 $\alpha_1 > \bar\kappa 2^{-c}\zeta^c/(1-\bar\gamma)$，然后选择 $\alpha_3 > \alpha_1 A_{13}/(1-A_{33})$。在这两种情况下，每个系数都小于 1；由于 $V'4, V'_5 \ge 0$且 $\mathrm{E}'V_5 \le L_5$，式 (8) 对某个 $\rho < 1$成立。 □

### 4.6 命题 1 的证明

所有分布 Gamma$(d/2, d/2)$，$d > 0$ 都与 $\mathbb{R}_+$ 上的勒贝格测度等价，且具有所有正矩；所有分布 Beta$(\nu, 1)$，$\nu > 0$ 都与 $(0, 1)$ 上的勒贝格测度等价；因此 (A0) 成立，并且对每个 $c$ 都有 $m_{c+1/2} < \infty$。$R_{H_d}(c)$ 的公式可由倾斜分布的伽马形式推出，而 slash 族的公式则由尾部参数为 $a$、在零点具有指标的分布的渐近式 $\mathrm{E}_s[\eta^{-c}] \sim \{\Gamma(a+\frac{1}{2}-c)/\Gamma(a+\frac{1}{2})\}s^c$ 得出（de la Cruz and Humeniy, 2026, 命题2(c̄)），其中 $a = \nu$。由于 $x \mapsto \Gamma(x-c)/\Gamma(x)$ 对 $c > 0$ 在 $(c, \infty)$ 上递减，$R_{H_d}(c)$ 随尾部参数递减，且网格上的最大值在 $d_{\min}$ 处取得。(a) 当 $c = 1$ 时，对于 Student-$t$ 分布有 $R_{H_d}(1) = 1/(d-1)$，对于 slash 分布有 $1/(2\nu - 1)$，且 $s_j(1) = \ell_j \le \ell_k^*$。(b) 若 $\delta_k = 0$，则 $\Pi_Z$ 中分块 $k$ 对应的列均为零，因此对于这些 $j$, $s_j(c) = 0$。(c) 若 $S_k \equiv 1$，则 (H2) 表示为 $R_{H_{d_{\min}}}(c) < 1$；由于 $R_H(0) = 1$，且当 $c = 0$ 时，log $R_H(c)$ 等于 $\psi(1\,2) - \psi(a + \frac{1}{2}) \prec 0$，其中 $\psi$ 是双伽马函数，且 $a = d_{\min}/2$ 或 $a = d_{\min}$，故对较小的 $c > 0$ 成立。最后一个结论由 de la Cruz and Humeniy (2026) 的命题 2(a)、(b) 和推论 1(iii)–(v) 得出。 □

### 4.7 命题 2的证明

论证沿用 de la Cruz 和 Humeniy (2026年，命题 3) 的证明，并且可原样应用于这三个模型，因为该证明只用到了 $\beta$ 的似然和先验分布，以及这些模型共有的精度参数。对于所有 $(\beta, u, \lambda_e)$，$p(y \mid \beta, u, \lambda_e) \le (\lambda_e/2\pi)^{N/2}$。因此，对于参数的任意非负函数 $F$，

$$\mathrm{E}_\pi[F] \le \frac{(2\pi)^{-N/2}}{m(y)}\,\mathrm{E}_{\text{prior}}[\lambda_e^{N/2} F],$$

其中 $m(y)$ 是边际似然。在先验分布下，$\beta, \lambda_e, \lambda_1, \dots, \lambda_K$ 相互独立，$\beta$服从正态分布，精度参数服从形状参数和率参数均为正的伽马分布。取 $F = \|\beta\|^m$, $F = |\log \lambda_e|^m$和 $F = |\log \lambda_k|^m$，右端有限，因为在这些先验分布下 $\mathrm{E}[\lambda_e^{N/2}]$、$\mathrm{E}\|\beta\|^m$、$\mathrm{E}[\lambda_e^{N/2}|\log \lambda_e|^m]$ 和 $\mathrm{E}|\log \lambda_k|^m$ 均有限。混合分布和尾部参数不起作用。

□

### 4.8 推论的证明1

令 $\tilde Z = ZC^{-1}$，并且对每个 $(i,\ d)$，设 $\tilde H_{i,\ d}$ 为 $\eta/c_i^2$ 在 $\eta \sim H_{i,\ d}$ 时的分布律。考虑模型 (1)–(3)˜ ˜，其中 $Z$ 和 $H_{i,\ d}$ 分别替换为 $Z$ 和 $H_{i,\ d}$，且所有先验分布保持不变，并考虑双射

$$\mathcal{T}(\lambda, d, \eta, \theta) = (\lambda, d, C^{-2}\eta, (\beta, Cu)).$$

若 $u_i \mid \lambda,\ \eta \sim \mathrm{N}(0,\ (\lambda_{k(i)}\eta_i)^{-1})$，则 $c_i u_i \mid \lambda,\ \eta \sim \mathrm{N}(0,\ (\lambda_{k(i)}\eta_i/c_i^2)^{-1})$，且 $\tilde Z(Cu) = Zu$。因此，在原模型下，$(\mathcal{T}(\lambda,\ d,\ \eta,\ \theta)\ Y)$ 的联合分布律就是参数的联合分布律以及

$Y$ 在重缩放模型下的分布律。因此，重缩放模型的后验分布是$\pi$ 在 T 下的像。由于链必须从任意起点出发都一致，我们直接检验：重缩放模型中每个完全条件分布在 T $(x)$ 处的取值，都是原模型中对应完全条件分布在 $x$处经相应坐标映射后的像。对于 $\lambda$，伽马参数取决于 $\|y-W\theta\|^2$和$\sum_i$: $k$ $(i)=k$ $\eta_i u_i^2$，二者均保持不变。对于 $(d,\eta)$，坐标 $i$ 的倾斜参数变为 $\lambda_{k(i)}c_i^2u_i^2/2 = c_i^2 s_i$，且 Z $\tilde{H}_{i,d}(c_i^2 s_i)$ $= \int(\eta/c_i^2)^{1/2}e^{-s_i\eta}H_{i,d}(d\eta)$ $= c_i^{-1}$Z$H_{i,d}(s_i)$；因子 $\prod_{i:k(i)=k} c_i^{-1}$不依赖于 $d_k$，并且在 (5) 中相消，因此 $d_k$ 的条件分布律保持不变；给定 $d_k$后，倾斜分布 ( $\tilde{H}_{i,d})c_i^2 s_i$ 是 $\eta/c_i^2$的分布律，其中 $\eta \sim (H_i,$ $d)s_i$。对于 $\theta$,~，条件分布律为高斯分布，且映射 $\theta \mapsto (\beta, Cu)$是线性的，其中 $W(\beta,Cu) = W\theta$ 且$\sum_i \lambda_{k(i)}(\eta_i/c_i^2)(c_iu_i)^2 = \sum_i \lambda_{k(i)}\eta_i u_i^2$，因此重缩放模型的条件分布律是~原模型条件分布律的像。因此，从 T $(x)$ 出发的重缩放模型 Gibbs 链 $\Phi$，其分布律与从 $x$出发的 T $(\Phi)$ 的分布律相同；对每个 $n$，均有 $\|P^n$(T $(x)$, $\cdot) - \tilde{\pi}\|_{\mathrm{TV}} = \|P_\Phi^n(x,\cdot) - \pi\|_{\mathrm{TV}}$，因为全变差距离在可测且逆映射也可测的双射下保持不变。因此，只需验证重缩放模型满足定理1的假设。

~ ~条件 (C1)–(C2) 保持不变，因为 rank$(Z)$ = rank$(Z)$ 且 rank$(Z_{-k})$ = rank$(Z_{-k})$。分布律 $\tilde{H}_{i,d}$, $d \in$ D$^{k(i)}$ 相互绝对连续，支撑为 $c_i^{-2}$S$_i$，并且 $m_r(\tilde{H}_{i,d})$ $= c_i^{-2r}m_r(H_{i,d})$，因此 (A0) 和 (H1) 中关于矩的部分成立。对于倾斜分布，$\tilde{H}_{i,d}$ 在 $s$ 处倾斜后是 $\eta/c_i^2$ 的分布律，其中 $\eta \sim (H_{i,d})^{s/c_i^2}$，因为 $x^{1/2}e^{-sx}$ 与 $x = \eta/c_i^2$ 成正比于 $\eta^{1/2}e^{-(s/c_i^2)\eta}$。因此 E$_{(\tilde{H}_{i,d})_s}[x^{-c}] = c_i^{2c}$ E$_{(H_{i,d})_{s/c_i^2}}[\eta^{-c}]$，并且

$$\kappa_{\tilde{H}_{i,d}}(c) = \limsup_{s\to\infty} \frac{c_i^{2c}\mathrm{E}_{(H_{i,d})_{s/c_i^2}}[\eta^{-c}]}{s^c} = \limsup_{t\to\infty} \frac{\mathrm{E}_{(H_{i,d})_t}[\eta^{-c}]}{t^c} = \kappa_{H_{i,d}}(c).$$

因此 $R_j(c)$ 保持不变，而投影到 $\mathcal{N}(\tilde{Z}) = C\mathcal{N}(Z)$ 得到 $s_j(c;\ C)$。条件 (H2)对于重缩放模型而言，就是该推论的假设。 □

### 4.9 推论 2 和 3 的证明

令 $m = q - r$，并将坐标按顺序排列，使得 $J = \{1,\dots,m\}$。由于 $Z$中位于 $J$之外的列线性无关，且张成 $Z$的列空间，因此对于每个 $v_J \in \mathbb{R}^m$，都恰好存在一个$v \in$ N $(Z)$，使其前 $m$ 个坐标满足条件。因此，N$(Z)$是矩阵 $\begin{pmatrix} I_m \\ A \end{pmatrix}$的列空间，其中$A$的阶数为 $(q-m) \times m$。对于 $M \geq 1$，令 $C_M =$ diag$(M1_m,$ $1_{q-m})$。那么 $C_M$N $(Z)$是 $B_M = \begin{pmatrix} I_m \\ A/M \end{pmatrix}$的列空间，其正交投影为 $\Pi_M = B_M(B_M^TB_M)^{-1}B_M^T$。由于 $M \to \infty$,$B_M^TB_M = I_m + A^TA/M^2 \to I_m$，因此 $\Pi_M \to$ diag$(1_m, 0_{q-m})$逐项成立。对于固定的 $c \in (0,1]$, $x \mapsto |x|^{2c}$是连续的，因此对于 $j \in J$，$s_j(c;\ C_M) \to 1$；对于 $j \notin J$，$s_j(c;\ C_M) \to 0$。如果对于 $j \in J$, $R_j(c) < 1$，那么一旦 $M$足够大，对于每个 $j$，都有$R_j(c)s_j(c;\ C_M) < 1$，而推论 1 得出推论 2。

对于推论3，任取一个主元集 $J$。对于 Student-$t$ 或 slash 分量，根据命题 1 及其证明中的导数计算，对所有充分小的 $c$ 和 $R_j(c) = R_{H_{d_{\min}}}(c) < 1$，当 $c > 0$ 充分小时，(H1) 均成立。对于其他分布族，$\kappa_H(c)$ $= 0$，因此 $R_j(c) = 0$，并且对于污染正态分布和正态-GIG 分布，(H1) 对所有 $c \in (0,1]$ 均成立；对于形状参数为 $\alpha$ 的方差-伽马分布，(H1) 对 $c < \alpha - \frac{1}{2}$ 成立（de la Cruz 和 Humeniy，2026，命题 2 和推论 1）。选取足够小的 $c$，使其同时满足所有分量的条件，再应用推论 2，即可证明该结论。 □

### 定理 2 和推论 4 的证明

我们首先给出高斯向量分数阶矩的一个界。

**引理 7.** 设 $\xi \sim \mathrm{N}_n(0,\Sigma)$ 且 $c \in (0,1]$。则 $\mathrm{E}\|\xi\|^{2c} \le \vartheta_n(c)(\operatorname{tr}\Sigma)^c$，其中 $\vartheta_n(c) \le 1$ 如第 *3.5* 节所述，且当 $\Sigma$ 是 $I_n$ 的倍数时等号成立。

证明。设 $\omega_1,\dots,\omega_n \ge 0$ 为 $\Sigma$ 的特征值，因此 $\|\xi\|^2$ 服从 $\sum_i \omega_i \zeta_i^2$ 的分布律，其中 $\zeta_1,\dots,\zeta_n$ 相互独立且服从标准正态分布。函数 $\phi(\omega) = \mathrm{E}(\sum_i \omega_i \zeta_i^2)^c$ 在 $\mathbb{R}_+^n$ 上是凹函数，因为 $x \mapsto x^c$ 是凹函数，且其自变量关于 $\omega$ 是线性的；同时该函数是对称的，因为 $\zeta_i$ 是可交换的。对 $\omega$ 的 $n!$ 个排列取平均并利用凹性，得到 $\phi(\omega) \le \phi(\bar\omega 1_n)$，其中 $\bar\omega = \operatorname{tr}\Sigma/n$。最后，$\phi(\bar\omega 1_n) = \bar\omega^c \mathrm{E}[(\chi^2_n)^c] = \bar\omega^c 2^c \Gamma(c+\tfrac{n}{2})/\Gamma(\tfrac{n}{2}) = \vartheta_n(c)(\operatorname{tr}\Sigma)^c$，并且根据詹森不等式，$\vartheta_n(c) \le 1$。 □

定理 2 的证明。下面说明定理 1 的证明需要作出的修改。引理 2 对于 $H_s^{(n)}$ 仍成立，只需将 $m_{1/2}$ 替换为 $m_{n/2}$：de la Cruz 和 Humeniy（2026，引理 5）中的证明仅用到 $H_s$ 的密度相对于固定有限测度正比于 $e^{-s\eta}$，此处该测度为 $\eta^{n/2}H(d\eta)$。特别地，$\mathrm{E}_s^{(n)}[\eta^c] \le m_{c+n/2}/m_{n/2}$。引理 4 的证明也相同，其中 $\bar H_b = H_b$、$d_b^\circ$、$h_b$、$d = dH_b$、$d/d\bar H_b$ 和 $\varphi = \bigotimes_b \bar H_b \otimes \mathrm{Leb}_{p+q}$，此时倾斜因子为 $h_b$、$d(\eta_b)\eta_b^{n_b/2} e^{-s_b'\eta_b}/\mathrm{Z}^{(n_b)}$ $H_b$、$d(s_b')$；对于固定的目标，它们与之前一样被某个常数乘以 $\exp(\sum_k \lambda_k b_k/2)$ 所支配，因为 $\mathrm{Z}^{(n_b)}$ $H_b$、$d(s) \ge e^{-s(\eta_b+\delta)} \int_{(0,\eta_b+\delta)} t^{n_b/2} H_b$、$d(dt) > 0$。引理 1、3、5 和 6 仅用到 $D_\eta$ 和 $\Lambda_\lambda$、$\eta$ 是正对角矩阵；当 $D_\eta = \operatorname{diag}(\eta_{b(1)},\dots,\eta_{b(q)})$ 时，这一点仍然成立，而且伽马全条件分布和式（10）保持不变。我们有以下内容取代式（9）：令 $s_b' = \lambda_{k(b)}\|u_b'\|^2/2$，

$$\mathrm{E}[\eta_b^c \mid \lambda,\theta'] \le M_c, \qquad \lambda_{k(b)}^{-c}\mathrm{E}[\eta_b^{-c} \mid \lambda,\theta'] \le \kappa_b' 2^{-c}\|u_b'\|^{2c} + C_0\lambda_{k(b)}^{-c}, \tag{11}$$

其中 $\kappa_b' = \max_d \kappa^{(n_b)}$ $H_b,\ d(c) + \epsilon$，$M_c = \max^{m_{c+n_b/2}(H_b,\ d)}/m_{n_b/2}(H_b,\ d)$ 以及 $C_0$ 都是对满足 $d \in \mathcal{D}^{k(b)}$ 的各对 $(b,\ d)$ 取最大值，而 $\epsilon > 0$ 的选取使得

$$\bar\gamma := (1+\epsilon)^2 \max_{b'} \kappa_{b'}' \frac{\Gamma(c+\frac{n_{b'}}{2})}{\Gamma(\frac{n_{b'}}{2})}\sigma_{b'}(c) < 1,$$

由于当 $\epsilon \to 0$ 时 $\bar\gamma \to \max_{b'} R_{b'}(c)\sigma_{b'}(c) < 1$，因此这是可能的。我们使用

$$v(\eta,\theta) = \alpha_1 \sum_{b\in\mathcal{B}} \|u_b\|^{2c} + \alpha_2 \|D_\eta^{1/2} u\|^{2c} + \alpha_3 \|y - W\theta\|^{2c} + \sum_{b\in\mathcal{B}} (\eta_b^c + \eta_b^{-c}) + \|\beta\|^{2c},$$

其中包含定理1证明中的 $V_5 = \|\beta\|^{2c}$ 项。根据定理1所采用的论证，其次水平集是紧的：在 $\{v \le \bar v\}$ 上，每个 $\eta_b$ 都属于紧集 $\mathsf{S}_b \cap [\bar v^{-1/c}, \bar v^{1/c}]$，$\|u_b\|$ 对 $i \in b$ 有界，且 $\|\beta\|^{2c}$ 有界，$\beta$ 无论 $X$ 的秩如何。项 $V_2$、$V_3$ 和 $V_5$ 的处理与之前相同。对于 $V_2$，分块内权重的重复不会影响界，因为 $\mathrm{E}[\sum_i \eta_{b(i)}^c \mid \lambda,\theta'] = \sum_b n_b \mathrm{E}[\eta_b^c \mid \lambda,\theta'] \le qM_c$，这不要求权重具有独立性。对于 $V_5$，引理3给出 $\|\mathrm{E}(\theta \mid \lambda,\ \eta)\|$ 和 $\mathrm{Var}(\beta \mid \lambda,\ \eta) \preceq \Sigma_\beta$ 的界，从而 $\mathrm{E}'V_5 \le L_5$。对于 $V_4$，(11) 和 $\bar{\mathrm{E}}'[\lambda_k^c] \le \zeta_k^c$ 给出 $\mathrm{E}'V_4 \le \bar\kappa 2^{-c}\zeta^c V_1' + L_4$。

对于 $V_1$，给定 $(\lambda,\eta)$，令 $u_b \sim \mathrm{N}_{n_b}(m_b,\Sigma_b)$。由引理 3，$\|m_b\| \le K_1$；记 $u_b = m_b + \xi_b$，利用第4节中的基本不等式和引理 7，可得 $\mathrm{E}\|u_b\|^{2c} \le (1+\epsilon)\vartheta_{n_b}(c)(\operatorname{tr}\Sigma_b)^c + C_{\epsilon,2c}K_1^{2c}$。结合 $\sigma_i^2 = (Q_\lambda^{-1},\eta)_{ii}$，定理1证明中使用的分解给出 $\sigma_i^2 \le (1+\epsilon)\sum_j \Pi_{ij}^2(\lambda_{k(j)}\eta_{b(j)})^{-1} + C_{\epsilon,2}c_\lambda^{-1}\{(Z^TZ)^+\}_{ii}$。对 $i \in b$ 求和，并按分块对 $j$ 分组，

$$\operatorname{tr}\Sigma_b \le (1+\epsilon)\sum_{b'} F_{bb'}\,(\lambda_{k(b')}\eta_{b'})^{-1} + C_{\epsilon,2}\tau_Z c_\lambda^{-1},$$

并且，由 $(\sum_j w_j)^c \le \sum_j w_j^c$ 和 $(1+\epsilon)^c \le 1+\epsilon$，

$$\mathrm{E}(V_1 \mid \lambda, \eta) \le (1+\epsilon)^2 \sum_{b'} \Big( \sum_{b} \vartheta_{n_b}(c) F_{bb'}^{c} \Big) \lambda_{k(b')}^{-c} \eta_{b'}^{-c} + C_8 c_\lambda^{-c} + C_9.$$

根据(11)，$\|u'_{b'}\|^{2c}$ 在 $\mathrm{E}(V_1 \mid \lambda, \theta')$ 中的系数为 $(1+\epsilon)^2 \kappa'_{b'} 2^{-c} \sum_b \vartheta_{n_b}(c) F_{bb'}^{c}$。由于 $2^{-c}\vartheta_{n_{b'}}(c) n_{b'}^{c} = \Gamma(c + \frac{n_{b'}}{2}) / \Gamma(\frac{n_{b'}}{2})$，这等于 $(1+\epsilon)^2 \kappa'_{b'} \Gamma(c + \frac{n_{b'}}{2}) \Gamma(\frac{n_{b'}}{2})^{-1} \sigma_{b'}(c) \le \bar{\gamma}$。如果 $\delta_{k(b')} = 0$，则 $\Pi_Z$ 中对应于 $b'$ 的所有列都为零，因此对于每个 $b$，分块 $b'$ 均不产生贡献。根据(11)，$\mathrm{E}(V_1 \mid \lambda, \theta')$的其余项因而至多为 $C \sum_{k \in \mathrm{K}_+} \lambda_k^{-c} + C_8 c_\lambda^{-c} + C_9$；而(10)仅在 $k \in \mathrm{K}_+$时使用，再结合 $c_\lambda^{-c} \le \lambda_e^{-c} + \varsigma_{\max}^{2c}$，可将它们的 $\mathrm{E}'$-期望界定为 $A_{12} V'_2 + A_{13} V'_3 + L_1$。因此，$\mathrm{E}' V_1 \le \bar{\gamma} V'_1 + A_{12} V'_2 + A_{13} V'_3 + L_1$，其中 $\bar{\gamma} < 1$，结论即如定理1的证明中那样得到，包括 $\mathrm{K}_+ = \emptyset$的情形。 □

推论4的证明。令 $\Psi_g^{1/2}$ 为对称平方根，并对 $v_b \mapsto \Psi_{g\,b}^{1/2}(\,) v_b$、$Z^{[b]} \mapsto Z^{[b]} \Psi_{g\,b}^{-1/2}(\,)$ 进行变换。与推论 1 的证明一样，该双射将模型映射到模型 (1)、(6)、(3)，其中分块大小为 $n_0$（由受试者给出），分量由各组给出，混合分布保持不变；它还将每个完全条件分布映射到相应的完全条件分布：伽马速率参数依赖于 $\eta_b v_b^T \Psi_{g(b)} v_b = \eta_b \|\Psi_{g\,b}^{1/2}(\,) v_b\|^2$，倾斜参数依赖于 $\lambda_{g(b)} \|\Psi_{g\,b}^{1/2}(\,) v_b\|^2/2$，而 $\theta$ 的更新服从高斯分布，且该映射是线性的。因此，两个 Gibbs 链以相同的速率收敛。该变换在右侧将 $Z$ 乘以一个可逆矩阵，该矩阵相对于各组呈分块对角结构，因此 (C1)–(C2) 中的秩保持不变。由于不同受试者对应的行互不重叠，变换后设计矩阵的零空间是各受试者对应的 $Z^{[b]} \Psi_{g\,b}^{-1/2}(\,)$ 的零空间的直和，因此 $\Pi_Z$ 是分块对角的，且对于 $b \ne b'$，$F_{bb'} = 0$。$\Pi_Z$ 的对角分块 $\Pi_b$ 是到 $\mathrm{N}(Z^{[b]} \Psi_{g\,b}^{-1/2}(\,))$ 上的正交投影，因此 $F_{bb} = \mathrm{tr}(\Pi_b^2) = \mathrm{tr}\ \Pi_b = \mathrm{nul}_b$。于是 $\sigma_b(c) = (\mathrm{nul}_b / n_0)^c$，定理 2 给出了第一个结论。

对于 $H_d =$ 伽马分布$(d/2, d/2)$，倾斜分布为伽马分布$(A, \frac{d}{2} + s)$，其中 $A = \frac{d+n_0}{2}$，因此 $\mathrm{E}^{(n_0)}_{s}[\eta^{-c}] = \Gamma(A-c)\Gamma(A)^{-1}(\frac{d}{2}+s)^c$。对于 slash分布 $H_\nu =$ 贝塔分布$(\nu, 1)$ 和 $A = \nu + \frac{n_0}{2}$，当 $s = 0$时，倾斜分布为贝塔分布$(A, 1)$；当$s > 0$时，倾斜分布为截断在 (0, 1) 上的伽马分布$(A, s)$，此时 $\mathrm{E}^{(n_0)}_{s}[\eta^{-c}] = s^c \Gamma(A-c) P(A-c, s) / \{\Gamma(A) P(A, s)\}$，因此 $\kappa^{(n_0)}_{H_\nu}(c) = \Gamma(A-c)/\Gamma(A)$，因为 $P(x, s) \to 1$ 如 $s \to \infty$。两种情况下，所有矩 $m_r$，$r > 0$ 均有限，并且 $\log \kappa^{(n_0)}_{H}(c) = \log \Gamma(c + \frac{n_0}{2}) - \log \Gamma(\frac{n_0}{2}) + \log \Gamma(A-c) - \log \Gamma(A)$ 在 $c = 0$ 处为零，导数$\psi(\frac{n_0}{2}) - \psi(A) < 0$也在该处成立，因为 $A > \frac{n_0}{2}$。因此，对于网格 $H$中的每个分布律，$\kappa^{(n_0)}_{H}(c) < 1$在区间 $(0, c_H)$ 上成立；又由于网格是有限的，存在某个 $c > 0$，使得对于所有受试者和尾部参数，均同时满足 $\kappa^{(n_0)}_{H}(c)(\mathrm{nul}_b/n_0)^c \le \kappa^{(n_0)}_{H}(c) < 1$。 □

## 5 插图 模型：$\beta$-HCG 数据的随机截距和斜率

我们使用 de la Cruz 和 Humeniy (2026) 分析的 $\beta$-HCG 数据：对 173 名前往生育诊所就诊的女性进行的 $N = 375$ 次 $\log_{10}\beta$-HCG 测量（Marshall and Barón，2000年），其中 124 人为正常妊娠，49 人出现异常结局。固定效应与 de la Cruz 和 Humeniy (2026) 中的相同：截距、以周为单位的时间 $t$、$t^2$、组指示变量及其与时间的交互项，因此 $p = 5$；$\beta_4$ 是两组在 $t = 0$时的差异。该论文发现，与具有共同尺度的正态随机截距模型相比，主要偏离之处在于异常妊娠的受试者间变异更大。这里，我们为每位女性加入时间随机斜率，并让四个分量（正常妊娠和异常妊娠的截距与斜率）各自具有独立的尺度和未知自由度。因此，$Z = (Z_1 \ Z_1 \circ t)$，其中 $Z_1$ 为受试者指示矩阵，$q = 346$分成 $K = 4$ 块。

表1： $\beta$-HCG 数据的随机截距和斜率：每个分量中进入定理 1 的各项，以及 Student-$t$ 随机效应自由度下界：取 $c = 1$ ($d > 1 + \ell_k^*$) 时，取每个分量的最佳 $c \in (0, 1]$ 时，以及采用示例 1 和 $c = \frac{1}{2}$ 中的对称重标度时。根据推论 3，实际上允许任意 $d > 0$。$L_k = \sum_{i:\ k(i)=k} \ell_{i_\circ}$

| 分量 | $q_k$ | $r_{-k}$ | $\delta_k$ | $L_k$ | $\ell_k^*$ | $c = 1$ | 最佳 $c$ | 对称 |
|---|---|---|---|---|---|---|---|---|
| 截距，正常 | 124 | 205 | 35 | 33.5 | 0.992 | 1.992 | 1.314 | 1 |
| 截距，异常 | 49 | 262 | 17 | 15.8 | 0.992 | 1.992 | 1.317 | 1 |
| 斜率，正常 | 124 | 205 | 35 | 1.5 | 0.109 | 1.109 | 0.197 | 1 |
| 斜率，异常 | 49 | 262 | 17 | 1.2 | 0.200 | 1.200 | 0.445 | 1 |

条件。由于 52 名女性只有单次测量， $r = \mathrm{rank}(Z) = 294$。表1列出了定理1中的各项数值。伽马形状参数满足条件 (C1)–(C2)，对任意正值均成立。仅有单次观测且观测时间为 $t$ 的女性，其截距具有 $\ell_i = t^2/(1 + t^2)$；由于这些时间范围为 2 至 11.4 周，截距分量的 $\ell_k^*$ 为 0.99。给定 $c = 1$，命题1(a)因此要求截距满足 $\min \mathrm{D}_k > 1.99$ 。对 $c$ 进行优化后，得到表1“最佳 $c$”列中较小的阈值；若所有分量共用一个 $c$ ，则要求 $\min \mathrm{D}_k > 1.32$。 $Z$ 的零空间由 52 对互不相交的元素组成，因此适用例1；采用 $c = \frac{1}{2}$ 进行对称重标度后，得到 $\min \mathrm{D}_k > 1$。最后，单次观测女性的 52 个截距构成一个主元集，推论3表明，对于任意有限网格，该链均为几何遍历。作为比较，若像 Román 和 Hobert（2015年，引理 3）的迹不等式那样，以秩亏度 $q - r$来界定迹的秩亏部分，则在 $c = 1$ 下得到 $d > 1 + q - r = 53$。

模型与先验分布。我们拟合了四个模型：M0，具有组特异尺度的正态随机截距，这是 de la Cruz and Humeniy (2026) 首选的模型；M1，具有四个尺度的正态随机截距和斜率；M2，采用相同设定，但随机效应服从 Student-$t$ 分布，且每个分量的自由度 $d_k$ 未知；以及 M3，Student-$t$ 分布的随机截距，具有两个组特异尺度，且 $d_k$ 未知。先验分布为 $\beta \sim \mathrm{N}_5(0, 1000I)$、$\lambda_e \sim$ 伽马分布(1, 0.1)、截距分量的 $\lambda_k \sim$ 伽马分布(1, 0.1)，以及斜率分量的 $\lambda_k \sim$ 伽马分布(1, $0.1/t^2$)，其中 $t^2 = 27.4$ 是各测量值中 $t^2$ 的均值。最后这一选择使得斜率的精度 $(t^2)^{1/2}$ 倍与截距的精度具有相同的先验分布，并使随机效应尺度的先验分布不随时间单位而改变。 $d_k$ 在网格 {2, 2.5, 3, 4, 5, 6, 8, 10, 15, 20, 30, 50, 100} 上具有相互独立的均匀先验分布，在此条件下 $\Pr(d_k \le 4) = 0.31$ ，且 $\Pr(d_k \le 10) = 0.62$。定理1和推论3给出了这四个模型的几何遍历性；对于 M2 和 M3，这也可由命题 1 在 $c = 1$ 条件下推出，因为 $\min \mathrm{D}_k = 2 > 1 + \ell_k^*$。

输出分析。对于每个模型，我们从分散的初始值出发运行了四条链，每条链迭代 30,000 次，并舍弃每条链的前 5,000 次迭代。我们监测了 $g = (\beta^T, \log\ \sigma_e, \log\ \sigma_1, \ldots, \log\ \sigma_K)^T$，其中$\sigma_e = \lambda_e^{-1/2}$ 且 $\sigma_k = \lambda_k^{-1/2}$ 。根据命题 2， $g$ 的所有阶后验矩均有限；几何遍历性保证了 $g$ 满足中心极限定理，且其渐近协方差的多元批均值估计量具有强一致性，批大小为 $\lfloor n^{0.6} \rfloor$。对于 M1 和 M2，每次迭代的多元有效样本量介于 0.29 和 0.37 之间；对于 M0 和 M3，则介于 0.50 和 0.78 之间。因此，每个模型的合并样本量都远远超过 M1 和 M2 的 $\mathrm{minESS}(10, 0.05, 0.05) = 8831$ ，其中 $g$有十个分量；也远远超过 M0 和 M3 的 $\mathrm{minESS}(8, 0.05, 0.05) = 8804$。 $g$ 中各分量的最大蒙特卡罗标准误为 0.013，八条半链的分裂潜在尺度缩减因子均不超过 1.0009。运行 M2 约需两分

表2： $\beta$-HCG数据：模型比较和后验汇总。M0：正态随机截距，两个尺度；M1：正态随机截距和斜率，四个尺度；M2：具有未知 $d_k$的Student-$t$随机截距和斜率，四个尺度；M3：具有未知 $d_k$ 的Student-$t$ 随机截距，两个尺度。括号内为标准误； ΔWAIC表示与M0的差值。 $\beta_4$ 为 $t = 0$时的组间差异， $\beta_4 + 4\beta_5$ 为 $t = 4$ 周时的组间差异；给出后验均值及95%区间。

| 模型 | WAIC | ΔWAIC | LPML | $\beta_4$ | $\beta_4 + 4\beta_5$ |
|---|---|---|---|---|---|
| M0 | 465.4 (41.4) | 0 | $-232.7$ | 0。20 ($-0.15$, 0。53) | $-0.49$ ($-0.70$, $-0.29$) |
| M1 | 471.5 (41.4) | 6.1 (4.6) | $-236.0$ | 0.24 ($-0.10$,0.58) | $-0.50$ ($-0.71$, $-0.29$) |
| M2 | 472.9 (41.5) | 7.5 (5.1) | $-236.5$ | 0。27 ($-0.08$, 0。60) | $-0.47$ ($-0.68$, $-0.25$) |
| M3 | 466.8 (41.6) | 1.4 (1.6) | $-233.4$ | 0.23 ($-0.12$,0.57) | $-0.46$ ($-0.67$, $-0.25$) |

表3： $\beta$-HCG 数据：自由度的后验分布 $d_k$。在网格上的均匀先验下，中位数为 8，概率($d_k \leq 4$) $= 0.31$ ，概率($d_k \leq 10$) $= 0.62$。通过对四条链每隔十次抽样取一次样本并计算批均值，得到后验概率的蒙特卡罗标准误，最大为 0.005。

| 模型 | 分量 | 后验中位数 | 概率($d_k \leq 4 \mid y$) | 概率($d_k \leq 10 \mid y$) |
|---|---|---|---|---|
| M2 | 截距，正常 | 20 | 0.04 | 0.31 |
| M2 | 截距，异常 | 10 | 0.21 | 0.54 |
| M2 | 斜率，正常 | 20 | 0.03 | 0.27 |
| M2 | 斜率，异常 | 8 | 0.28 | 0.61 |
| M3 | 截距，正常 | 20 | 0.04 | 0.31 |
| M3 | 截距，异常 | 10 | 0.20 | 0.53 |

半钟。

**模型比较**。 将受试者的随机效应和权重积分消去后，该女性测量值的似然是关于权重的正态密度混合；我们使用高斯–拉盖尔求积计算该似然，每个权重取 40 个求积节点，其结果与自适应积分一致，相对误差低于 $2 \times 10^{-6}$。据此，我们以受试者为单位计算了 WAIC 和 LPML，并根据受试者层面的贡献计算了 WAIC 的标准误及 WAIC 差值的标准误（Vehtari 等人，2017年）。

结果。表 2 和表 3 总结了拟合结果。WAIC 和 LPML 得出的排序相同。两个准则都没有提供支持随机斜率的证据：M1 的表现比 M0 差 6.1 个 WAIC 单位，标准误为 4.6；在 M1 下，斜率的尺度后验均值 $\sigma_k$ 较小（正常妊娠和异常妊娠每周分别为 0.03 和 0.08，而截距分别为 0.17 和 0.54）。它们也没有提供支持重尾的证据：将正态随机效应替换为 Student-$t$随机效应后，截距模型的 WAIC 增加了 1.4（标准误为 1.6），截距和斜率模型的 WAIC 增加了 1.4（标准误为 1.1）。因此，对于女性人数较多的分量，其 $d_k$ 后验分布向较大的取值移动；而对于异常妊娠的截距及其斜率（这些参数由 49 名女性提供信息，其中许多人只有一两次测量），后验分布则仍接近先验分布。这些结果与 de la Cruz 和 Humeniy (2026) 的发现一致：相较于共同尺度正态模型，主要差异在于异常妊娠的受试者间变异更大。关于组间差异的结论不受模型影响：在全部四个模型中， $t = 0$ 处的差异都不确定，概率($\beta_4 > 0 \mid y$)介于 0.87 和 0.94 之间；在四周时，异常妊娠组低约

0.5 在对数10 尺度上，且区间不包含零。每个 $d_k$ 的后验分布都是采样器的副产品，其蒙特卡罗误差也受同一理论控制，因为指示函数 $\mathbf{1}\{d_k \le d_0\}$ 有界，并且在几何遍历性条件下，中心极限定理适用于这些函数（Jones，2004）。

**对尺度先验分布的敏感性**。观察一次或两次的女性人数如此之多，斜率尺度的辨识较弱，因此其先验分布会产生影响。对所有精度参数采用 Gamma(1,0.1) 分布时，其密度在精度较大时快速衰减，斜率尺度的后验均值被推高至 0.07 和 0.15；M0、M1、M2 和 M3 的 WAIC 值分别为 465.4、514.5、516.7 和 466.8；M1 比 M0 差 49.1 个单位（标准误为 12.3）。对所有精度参数采用 Gamma(1,0.001) 分布时，正常妊娠截距的尺度后验均值从 0.17 降至 0.07，WAIC 值分别为 465.6、472.8、473.9 和 467.4；链的混合也有所恶化，截距模型和斜率模型每次迭代的多元有效样本量约为 0.15，分裂潜在尺度缩减因子最高达 1.05；因此，与其他先验相比，该先验下的结果精度较低。在三种先验下，模型排名和四周时的组间差异均相同，所有拟合中的 $\beta_4 + 4\beta_5$ 后验均值介于 $-0.50$ 和 $-0.45$ 之间。 $t = 0$ 处的差异则更为敏感，这符合外推的预期：在不同模型和先验下，$\Pr(\beta_4 > 0 \mid y)$ 的范围为 0.86 至 0.98。三种先验下均满足几何遍历性，但这并未对采样器的效率给出定量界限；在这里，先验分布非常弥散时，采样器效率会下降。

## 6　共享权重 ： 尺度与尾部参数的探索

当权重共享且尾部非常厚重时，通过蒙特卡洛能够多么精确地估计随机效应的尺度和尾部参数？为在推论4所涵盖的设定下考察这一问题，我们模拟了 $m = 60$ 个受试者，其中15人只有一次观测，其余受试者有两至六次观测，来自 $y_{ij} = \beta_1 + \beta_2 t_{ij} + v_{i1} + v_{i2} t_{ij} + e_{ij}$，其中 $t_{ij}$在(0, 1)上服从均匀分布，$\beta = (1, 0.5)^T$、 $\sigma_e = 0.3$ 以及 $v_i \mid \eta_i \sim \mathrm{N}_2(\mathbf{0}, 0.25\,\eta_i^{-1} I)$， $\eta_i \sim$伽马分布$(\tfrac{1}{2}, \tfrac{1}{2})$，因此随机效应向量服从二元柯西分布。因此 $N = 201$， $q = 120$ 且秩$(Z) = 105$，每个只有一次观测的受试者都携带零空间的一个一维部分（$\mathrm{nul}_b / n_0 = \tfrac{1}{2}$ ）。我们拟合了具有一个组的模型（7）-， $\Psi = I$，对 $d$ 在{0.5, 1, 2, 4, 8, 16}上采用均匀先验，对 $\beta \sim \mathrm{N}_2(\mathbf{0}, 1000 I)$以及两个精度参数均采用伽马分布(1, 0.1)先验。根据推论4，吉布斯采样器是几何遍历的。我们从分散的初始值出发运行了4条链，每条链迭代105,000次，并舍弃每条链的前5000次迭代。

$d$ 的后验分布集中在网格中两个最大的值上，其中 $\Pr(d = 0.5 \mid y) = 0.355$，$\Pr(d = 1 \mid y) = 0.645$。 $g = (\beta_1$、 $\beta_2$、log $\sigma_e$、log $\sigma_u)^T$的分裂 $\hat{}R$ 统计量最大为 1。0006。 $\beta_1$、 $\beta_2$ 和 $\log \sigma_e$的每次迭代有效样本量介于 0。26 和 0。41 之间，但 $\log \sigma_u$ 仅为 0。021，指示变量 $\mathbf{1}\{d = 0.5\}$仅为 0。030。$g$ 的每次迭代多元有效样本量约为 0。2，因此，按照 $8431/(\mathrm{mESS}/n)$ 外推，单条链大约需要 43，000 次迭代才能达到 minESS(4， 0。05，0。05) $= 8431$；四条链中 $\Pr(d = 0.5 \mid y)$ 的蒙特卡罗标准误为 0。004。

多元有效样本量与$\log \sigma_u$ 以及尾部指示变量的有效样本量之间的差异，促使我们开展一项小型实验，考察这些量的探索程度如何取决于每个受试者的信息量。在模型和网格相同的情况下，我们改变了具有单次观测的 $n_1 \in \{0、15、30\}$名受试者（共60名）的数量，使得秩$(Z) = 120 - n_1$，并改变真实自由度 $d_0 \in \{1、4\}$。对于六种配置中的每一种，我们生成了三个数据集，并运行两条链，各进行55,000 次迭代，舍弃前5000次。表4 报告了结果。推论4适用

表4：共享权重随机截距和斜率，60名受试者，其中 $n_1$ 名仅有单次观测；真实自由度为 $d_0$， $d$的均匀先验分布取值为{0.5、1、2、4、8、16}；每行对应三个数据集，每个数据集在预热期后运行两条链，各进行50,000次迭代。每次迭代的有效样本量：在各数据集及 $\beta_1$、 $\beta_2$ 上的回归系数取值范围；log $\sigma_e$ 和log $\sigma_u$在各数据集上的平均值；mESS/$n$ 对应 $g = (\beta_1$、 $\beta_2$、log $\sigma_e$、log $\sigma_u)^T$；Pr($d \le 1 \mid y$)在各数据集上的取值范围；对于该概率落在(0.02、0.98)内的数据集，$\mathbf{1}\{d \le 1\}$ 的每次迭代有效样本量；以及该概率的最大蒙特卡罗标准误和 $g$的最大分段 $R$ 。

| | | | 每次迭代的有效样本量 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $d_0$ | $n_1$秩($Z$) | | $\beta$ | log $\sigma_e$ | log $\sigma_u$ | mESS/$n$ | 概率($d \le 1 \mid y$) | ESS $\mathbf{1}\{d \le 1\}$ | MCSE | 最大值 $\hat{R}$ |
| 1 | 0 | 120 | 0.33–0.55 | 0.35 | 0.057 | 0.26 | 0.18–0.99 | 0.06–0.13 | 0.006 | 1.0004 |
| 1 | 15 | 105 | 0.29–0.65 | 0.34 | 0.042 | 0.24 | 0.95–1.00 | 0.18 | 0.002 | 1.0009 |
| 1 | 30 | 90 | 0.34–0.64 | 0.29 | 0.049 | 0.23 | 0.92–1.00 | 0.08 | 0.003 | 1.0013 |
| 4 | 0 | 120 | 0.47–0.82 | 0.36 | 0.092 | 0.36 | 0.00–0.01 | – | 0.001 | 1.0004 |
| 4 | 15 | 105 | 0.46–0.74 | 0.30 | 0.084 | 0.33 | 0.00 | – | 0.001 | 1.0003 |
| 4 | 30 | 90 | 0.47–0.79 | 0.28 | 0.068 | 0.31 | 0.00–0.01 | – | 0.001 | 1.0001 |

对它们所有人而言。

有三个特点值得注意。首先，在每种配置下，随机效应的尺度混合速度都远慢于回归系数：其每次迭代的有效样本量为 0.04 至 0.09，而 $\beta$为 0.29 至 0.82；尾部越厚，该有效样本量越低。 $g$ 的多元有效样本量每次迭代介于 0.23 至 0.36 之间，却未能体现这一点：它通过置信区域的体积概括联合精度，而多元准则并不能保证每个分量都达到规定的蒙特卡罗精度；它也不涉及尾部概率，因为尾部概率的指示变量不属于 $g$。其次，在这些模拟结果中，不同量之间的差异大于不同配置之间的差异：对于 $d_0 = 4$ ，当 $n_1$ 从 0 增至 30 时，log $\sigma_u$ 的有效样本量从 0.092 降至 0.068；但对于 $d_0 = 1$ ，该数值并未单调变化。每种配置下有三个数据集，其中信息量、秩和数据都同时发生变化，因此这项实验无法支持关于每个受试者的观测数所产生影响的一般性结论。第三，对尾部的探索精度取决于 $d$ 的后验分布有多不确定。对于 Pr($d \le 1 \mid y$) 的中间取值，指示变量每次迭代的有效样本量可能低至 0.06；即便如此，100,000 次抽样得到的该概率的蒙特卡罗标准误估计仍低于 0.006。当概率接近 0 或 1 时，其标准误估计较小，但这仅涉及事件 $\{d \le 1\}$ ，并不表明网格中的其他取值都得到了充分探索。

对于贝叶斯实践，理论和这些结果导向一项简单的建议。由于马尔可夫链是几何遍历的，且涉及的量要么有界，要么具有任意阶的后验矩，因此回归系数、尺度的对数以及尾部参数的后验概率均可计算有效的蒙特卡罗标准误；应分别计算并报告这些量的标准误。对于回归系数而言足够长的运行，或对于不针对每个分量设定目标的多元准则而言足够长的运行，对于随机效应的尺度以及有关尾部的判断来说，可能仍短了好几倍；而这些往往正是稳健分析所关注的量。

## 7 讨论

我们将针对具有正态随机效应尺度混合的线性混合模型的分块吉布斯采样器分析扩展到了多个分量、秩亏设计、有限网格上的未知尾部参数以及共享权重。新增的要点包括通过 $\delta_k$ 和引理5进行分量层面的秩亏控制，以及定理2中的分块论证；对设计矩阵进行重新缩放，是完善常见分布族结果的工具，此处采用了de la Cruz 和 Humeniy(2026年)针对这一情形所作的调整。关于混合分布的条件通过设计矩阵零空间上的投影来表述，该投影可在运行采样器之前计算；推论1 和2 表明，对设计矩阵的列重新缩放后即可检验这些条件。采用个体权重时，适当的重新缩放可使所有常见分布族和任意设计均满足这些条件，因此设计矩阵的秩仅通过(C1) 和(C2)限制精度参数伽马先验的形状参数。采用共享权重时，定理2仍要求零空间的分块满足几何条件；对于推论4中的个体特异性设计，可得到适用于任意有限网格上 Student-$t$ 和 slash 尾部参数的结果。

三个扩展方向似乎是自然而然的。首先，推论4涵盖了具有已知形状矩阵和未知尺度的多元尺度混合分布。对于具有未知非结构化精度矩阵 Ξ以及尺度矩阵为 $S_0$的 Wishart先验的情形，采样器仍保持其三分块结构，但漂移论证无法延续。在 $Z$的零空间方向上，某个受试者随机效应的条件方差取决于该方向上的 Ξ ，而共享权重则受到当前随机效应沿 Ξ 的二次型的倾斜；二者不再像定理1的证明中那样相互抵消尺度，一个步骤的界会产生漂移函数各项的乘积。对于误差为正常尺度混合、共享权重结构但不存在未识别方向的多元回归，Roy 和 Hobert（2010年）以及 Hobert 等人（2018年）在当前尺度矩阵所定义的度量下测量残差，此时相关项会相互抵消。这里的类似度量将通过共享权重和当前随机效应依赖于链的状态，而难点在于控制该度量从一次扫描到下一次扫描的变化。其次，对尾部参数使用连续先验会破坏尾部参数与权重联合更新的条件共轭性，并可能需要额外的非共轭更新，例如 Metropolis–Hastings 步骤；Jones 等人（2014年）给出了条件，使得条件 Metropolis–Hastings 采样器能够从相应的精确吉布斯采样器继承几何遍历性。我们的采样器先抽样 $\lambda$ ，然后从其完全条件分布中抽样（$d$， $\eta$），这通常并不等同于从给定 $\theta$时（$\lambda$， $d$， $\eta$）的联合条件分布中精确抽样，因此定理1本身无法提供那里所需的性质；建立该性质，或针对三分块结构建立类似结果，仍是一个开放问题。第三，重尾误差在实践中很常见，观测值带有潜在权重，而我们的实现支持这种情形；但它们会在 $\theta$的条件方差中引入第二组权重，而当前论证无法控制这一点。我们的条件是充分而非必要的。特别是，引理 5的使用中 $\Lambda$ 取决于权重，而利用 $D_\eta^{1/2}$N（$Z$）特定结构的界或许可以弱化（C2），尽管该条件在我们的应用中并不构成约束。

## 致谢

第一作者的工作部分得到了智利 ANID 的资助，资助项目包括 ANID/ FONDECYT/1260228、ANID/ANILLO/ACT210096 和 ANID-MILENIO-NCN2024_103，以及 Data Observatory Foundation 和 ANID 技术中心 DO210001 的资助。第二作者的工作部分得到了 ANID 博士奖学金的资助。

## 代码和数据可用性

包含吉布斯采样器的 Python 代码（该采样器具有多个分量、有限网格上的尾部参数和共享权重）、附录A的验证测试、第5节的分析以及第6 节的模拟均可在 https://github.com/rolandodelacruz74/smnlmm-components获取。$\beta$-HCG 数据可应要求向第一作者索取。

## 实现的验证

该采样器以 Python 实现，支持任意数量的分量、各分量特定的混合族、取值于有限网格上的尾部参数（与权重联合抽样），以及精确抽取 $\theta$（当 $Z$ 的每一行至多有一个非零元素时，通过舒尔补实现；否则通过对$P_\lambda$、$\eta$ 进行乔列斯基分解实现）。我们使用自适应求积对倾斜分布和归一化函数 $Z_{H_d}$ 进行了比较。我们还将尾部参数与权重的联合抽样结果同网格上的精确条件分布进行了核验。此外，我们采用五种配置、每种配置含 $N = 9$ 个观测值，运行了 Geweke（2004年）联合分布检验：四名受试者的随机截距和随机斜率，其中一名受试者只有一次观测（因此 $Z$ 秩亏），随机截距服从 Student-$t$分布， $d$ 未知且取值于{2.5, 4, 8}，随机斜率服从 Student-$t$ 分布， $d = 3$；slash 随机截距的 $\nu$ 未知，取值于{0.8, 1.5, 3}和{1.5, 3, 6}，并结合自由度未知的 Student-$t$误差项；Student-$t$ 随机截距与污染正态误差；以及带有 $\nu = 0.8$的 slash 随机截距。含重尾误差项的配置用于检验本研究设定之外的部分实现。所有精度参数的先验分布为 $\beta \sim \mathrm{N}_2(0, I)$和 Gamma(3, 2)。对于每种配置，我们在 60,000 次迭代中比较了 $\beta_1$、 $\beta_2$、 $\lambda_e$的均值、首个和末个分量的精度参数、 $|u_1|^{1/2}$、 $|u_q|^{1/2}$、各类潜在权重的对数，以及适用情况下的尾部参数，共进行了 48 项比较。我们没有比较随机效应本身，因为对于带有 $\nu = 0.8$ 的 slash 分布，其先验方差为无穷大；在所有配置中， $|u_i|^{1/2}$ 的先验方差均有限，因为对于所用的每种混合分布，$\mathrm{E}[\eta^{-1/2}] < \infty$ 且 $\mathrm{E}[\lambda_k^{-1/2}] < \infty$。以批均值标准误进行标准化（该标准误为此次经验检验提供了实用尺度）后，48 个差值中绝对值最大的为 2.99。代码可在 https://github.com/rolandodelacruz74/smnlmm-components获取。

## B 引理证明 1–3 及基本事实的证明

这些证明出自 de la Cruz 和 Humeniy (2026年)，此处采用本文的记号。

引理 1. (a) 由于 $Q \succeq \Omega \succeq \Upsilon$, $Q^{-1} \preceq \Omega^{-1} \preceq \Upsilon^{-1}$，可得第二个界。对于第一个界，只需证明 $P_A\Omega^{-1}P_A \preceq \upsilon^{-1}(A^TA)^+$，因为 $Q^{-1} \preceq \Omega^{-1}$。令 $k =$ 为$A$的秩。若 $k = 0$，则$P_A = 0$；若 $k = n$，则 $P_A = I$ 且 $\Omega \succeq \upsilon A^TA$，后者可逆。否则，令 $O = (O_1\ O_2)$ 为一个正交矩阵，其前 $k$ 列张成 $A$ 的行空间，后 $n-k$ 列张成 $\mathrm{N}(A)$，从而 $P_A = O_1O_1^T$且 $(A^TA)^+ = O_1B_A^{-1}O_1^T$ 为正定矩阵。取 $\Upsilon_{ab} = O_a^T\Upsilon O_b$,

$$O^T\Omega O = \begin{pmatrix} \upsilon B_A + \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{pmatrix},$$

分块矩阵的逆的公式给出 $O_1^T\Omega^{-1}O_1 = (\upsilon B_A + \Upsilon_{11} - \Upsilon_{12}\Upsilon_{22}^{-1}\Upsilon_{21})^{-1}$。正定矩阵 $O^T\Upsilon O$ 的舒尔补 $\Upsilon_{11} - \Upsilon_{12}\Upsilon_{22}^{-1}\Upsilon_{21}$ 是正定的，因此 $O_1^T\Omega^{-1}O_1 \preceq (\upsilon B_A)^{-1}$ 且 $P_A\Omega^{-1}P_A = O_1(O_1^T\Omega^{-1}O_1)O_1^T \preceq \upsilon^{-1}(A^TA)^+$。(b) 由于 $P_A + \Pi_A = I$且 $\mathrm{tr}(P_A\Omega^{-1}\Pi_A) = \mathrm{tr}(\Omega^{-1}\Pi_AP_A) = 0$, $\mathrm{tr}(\Omega^{-1}) = \mathrm{tr}(P_A\Omega^{-1}P_A) + \mathrm{tr}(\Pi_A\Omega^{-1}\Pi_A)$，而(a)部分

将这两项界定为 $\upsilon^{-1}\,\mathrm{tr}\{(A^TA)^+\}$ 和 $\mathrm{tr}(\Pi_A\Upsilon^{-1}\Pi_A)=\sum_i\tilde{\ell}_i\upsilon_i^{-1}$，其中最后一个等式使用了 $\Pi_A^2=\Pi_A$。(c) 令 $\upsilon_{\min}=\min_i\upsilon_i$，并令 $\vartheta_1,\ldots,\vartheta_n$ 为 $A^TA$的特征值。由于 $\Omega\succeq\upsilon A^TA+\upsilon_{\min}I$,

$$\mathrm{tr}(A\Omega^{-1}A^T)\le\mathrm{tr}\{A(\upsilon A^TA+\upsilon_{\min}I)^{-1}A^T\}=\sum_{i:\vartheta_i>0}\frac{\vartheta_i}{\upsilon\vartheta_i+\upsilon_{\min}}\le\frac{\mathrm{rank}(A)}{\upsilon}.$$

引理2. 若 $0\le s<s'$， $H_{s'}$ 关于 $H_s$ 的密度为 $w/\mathrm{E}_s[w]$，其中 $w(\eta)=e^{-(s'-s)\eta}$，且单调不增。对于有界的单调不减 $g$,

$$\mathrm{E}_{s'}[g]-\mathrm{E}_s[g]=\frac{\mathrm{E}_s[gw]-\mathrm{E}_s[g]\mathrm{E}_s[w]}{\mathrm{E}_s[w]}\le 0,$$

因为同一随机变量的单调不减函数与单调不增函数的协方差非正（切比雪夫关联不等式）。由单调收敛定理，对 $\min(g,n)$ 应用此结论，可知它对任意单调不减的 $g\ge 0$ 都成立，而对任意单调不增的 $g\ge 0$，反向不等式成立。(a) 令 $g(\eta)=\eta^c$，$\sup_{s\ge 0}\mathrm{E}_s[\eta^c]=\mathrm{E}_0[\eta^c]=m_{c+1/2}(H)/m_{1/2}(H)$。(b) 选取 $S>0$，使得对 $s\ge S$ 有 $\mathrm{E}_s[\eta^{-c}]\le\kappa' s^c$；根据 $\kappa_H(c)$ 的定义（作为上极限），这样的选择是可行的。由于 $\eta^{-c}$ 单调不增，$\mathrm{E}_s[\eta^{-c}]\le\mathrm{E}_S[\eta^{-c}]\le\kappa' S^c=:C$，对 $0\le s<S$ 成立。

引理3。由于 $\theta\mid\lambda,\ \eta$ 为正态分布，其均值使

$$\lambda_e\|y-W\theta\|^2+u^T\Lambda_{\lambda,\eta}u+(\beta-\mu_\beta)^T\Sigma_\beta^{-1}(\beta-\mu_\beta)=\|D^{1/2}(\tilde{y}-\tilde{A}\theta)\|^2,$$

其中

$$\tilde{A}=\begin{pmatrix}X & Z\\ 0 & I_q\\ \Sigma_\beta^{-1/2} & 0\end{pmatrix},\qquad \tilde{y}=\begin{pmatrix}y\\ 0\\ \Sigma_\beta^{-1/2}\mu_\beta\end{pmatrix},\qquad D=\mathrm{diag}(\lambda_e 1_N,\Lambda_{\lambda,\eta},1_p).$$

~ ~ $A$ 的最后 $q+p$ 行构成一个可逆矩阵，因此 $A$ 满列秩，且不依赖于（$\lambda$，$\eta$），而 $D$ 是一个正对角矩阵。因此，$\mathrm{E}(\theta\mid\lambda,\ \eta)=(\tilde{A}^TD\tilde{A})^{-1}\tilde{A}^TD\tilde{y}=\tilde{A}^+\{\tilde{A}(\tilde{A}^TD\tilde{A})^{-1}\tilde{A}^TD\}\tilde{y}$，其中 $\tilde{A}^+=(\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T$。根据 Stewart（1989）和 O'Leary（1990）定理，$\chi=\sup_D\|\tilde{A}(\tilde{A}^TD\tilde{A})^{-1}\tilde{A}^TD\|_2<\infty$，其中上确界取遍所有正对角矩阵$D$。因此，$\|\mathrm{E}(\theta\mid\lambda,\ \eta)\|\le K_1:=\|\tilde{A}^+\|_2\chi\|\tilde{y}\|$，且 $\|y-W\mathrm{E}(\theta\mid\lambda,\ \eta)\|\le K_2:=\|y\|+\|W\|_2K_1$。该论证仅用到$\Lambda_\lambda$，$_\eta$ 是一个正对角矩阵，因此相同的常数也适用于具有共享权重的模型。

基本事实。对于 $p\le 1$，$t\mapsto t^p$ 是次可加的，因此可以取 $C_{\epsilon,\ p}=1$；对于 $p>1$ 和 $t\in$（0，1），由凸性可得 $(x+y)^p=\{(1-t)\frac{x}{1-t}+t\frac{y}{t}\}^p\le(1-t)^{1-p}x^p+t^{1-p}y^p$，并将 $t$ 取为满足 $(1-t)^{1-p}\le 1+\epsilon$。若 $Y\sim\mathrm{N}(m,\ \sigma^2)$，则写作 $Y=m+\sigma\xi$，其中 $\xi\sim\mathrm{N}(0,\ 1)$；利用 $p=2c$，由第一个事实可得 $|Y|^{2c}\le(1+\epsilon)\sigma^{2c}|\xi|^{2c}+C_{\epsilon,\ 2c}|m|^{2c}$，取期望即可得到该界。不等式 $(\sum_j w_j)^c\le\sum_j w_j^c$ 是 $t\mapsto t^c$ 的次可加性；由于 $t\mapsto t^c$ 是凹函数，因此可以应用詹森不等式。

## 参考文献


Tavis Abrahamsen 和 James P. Hobert。带有 $p>N$ 的贝叶斯线性混合模型的分块吉布斯采样器收敛性分析。*Bernoulli*，23(1):459–478，2017年。doi: 10.3150/15-BEJ749。

Arka Banerjee 和 Dootika Vats。马尔可夫链蒙特卡罗中的多元强不变原理。《Electronic Journal of Statistics》，18(1)，2024年。doi：10.1214/24-EJS2257。

Rolando de la Cruz 和 Nathalie Humeniy。具有正态随机效应尺度混合的贝叶斯线性混合模型的分块吉布斯采样器的几何遍历性。arXiv:2611.05600，2026年。James M. Flegal、Murali Haran 和 Galin L. Jones。马尔可夫链蒙特卡罗：我们能相信第三位有效数字吗？《统计科学》，23：250–260，2008年。Thais C. O. Fonseca、Marco A. R. Ferreira 和 Helio S. Migon。Student-t回归模型的客观贝叶斯分析。《生物统计学》，95(2)：325–333，2008年。John Geweke。做好它：后验模拟器的联合分布检验。《美国统计学会会刊》，99：799–804，2004年。James P. Hobert、Yeun Ji Jung、Kshitij Khare 和 Qian Qin。具有非高斯误差的贝叶斯多元线性回归的 MCMC算法收敛性分析。《斯堪的纳维亚统计学杂志》，45(3)：513–533，2018年。doi：10.1111/sjos.12310。Galin L. Jones。关于马尔可夫链中心极限定理。《Probability Surveys》，1：299–320，2004年。Galin L. Jones、Gareth O. Roberts 和 Jeffrey S. Rosenthal。条件 Metropolis–Hastings 采样器的收敛性。《应用概率进展》，46：422–445，2014年。Victor H. Lachos、Pulak Ghosh 和 Reinaldo B. Arellano-Valle。偏斜正态独立线性混合模型的似然推断。《中华统计学志》，20(1)：303–322，2010年。Kenneth L. Lange、Roderick J. A. Little 和 Jeremy M. G. Taylor。使用 $t$ 分布的稳健统计建模。《美国统计学会会刊》，84(408)：881–896，1989年。doi：10.1080/01621459.1989.10478852。Guillermo Marshall 和 Anna E. Barón。不平衡纵向数据的线性判别模型。《医学统计学》，19(15)：1969–1981，2000年。doi：10.1002/1097-0258(20000815)19:15<1969::AID-SIM515>3.0.CO;2-Y。S. P. Meyn 和 R. L. Tweedie。《马尔可夫链与随机稳定性》。施普林格出版社，伦敦，1993年。Dianne P. O’Leary。关于缩放投影和伪逆的界。《线性代数及其应用》，132：115–117，1990年。José C. Pinheiro、Chuanhai Liu 和 Ying Nian Wu。使用多元 $t$ 分布在线性混合效应模型中进行稳健估计的高效算法。《计算与图形统计学杂志》，10(2)：249–276，2001年。doi：10.1198/10618600152628059。Jorge C. Román 和 James P. Hobert。具有非正规先验的贝叶斯一般线性混合模型的吉布斯采样器收敛性分析。《统计学年鉴》，40：2823–2849，2012年。Jorge C. Román 和 James P. Hobert。具有正规先验的贝叶斯一般线性混合模型的吉布斯采样器的几何遍历性。《线性代数及其应用》，473：54–77，2015年。Guilherme J. M. Rosa、Carlos R. Padovani 和 Daniel Gianola。具有正态/独立分布的稳健线性混合模型及其贝叶斯MCMC实现。《生物计量学杂志》，45 (5)：573–590，2003年。doi：10.1002/bimj.200390034。

Vivekananda Roy 和 James P. Hobert。关于具有重尾误差项的贝叶斯多元回归模型的蒙特卡罗方法。《多元分析杂志》，101(5):1190–1202，2010年。G. W. Stewart。关于缩放投影和伪逆。《线性代数及其应用》，112: 189–193，1989年。Luke Tierney。用于探索后验分布的马尔可夫链（附讨论）。《统计学年鉴》，22:1701–1762，1994年。Dootika Vats、J. M. Flegal 和 Galin L. Jones。马尔可夫链蒙特卡罗的多元输出分析。《生物统计学》，106(2):321–337，2019年。doi: 10.1093/biomet/asz002。Aki Vehtari、Andrew Gelman 和 Jonah Gabry。使用留一交叉验证和 WAIC 进行实用的贝叶斯模型评估。《统计与计算》，27(5):1413–1432，2017年。Cristiano Villa 和 Stephen G. Walker。关于 t 分布自由度的客观先验。《贝叶斯分析》，9(1):197–220，2014年。