arXiv:2602.03613v1 [stat.ME] 3 Feb 2026

通过回归投影和批次差异进行基于模拟的推断

Arya Farahi
ARYA.FARAHI@AUSTIN.UTEXAS.EDU德克萨斯大学奥斯汀分校统计与数据科学系，美国德克萨斯州奥斯汀，邮编 *78712 NSF*-西蒙斯宇宙起源人工智能研究所，美国

Jonah Rose JR8952@PRINCETON.EDU
普林斯顿大学物理系，美国新泽西州普林斯顿，邮编 *08544*；*Flatiron*研究所计算天体物理中心，美国纽约州纽约，邮编 *10010*

Paul Torrey PAUL.TORREY@VIRGINIA.EDU
弗吉尼亚大学天文学系，美国弗吉尼亚州夏洛茨维尔，邮编 *22904*；弗吉尼亚大学理论天文学研究所，美国弗吉尼亚州夏洛茨维尔，邮编 *22904*；美国国家科学基金会—西蒙斯宇宙起源人工智能研究所

摘要

我们分析了一种轻量级的基于模拟的推断方法，该方法仅利用对观测数据进行回归投影来推断模拟器参数。拟合一次代理线性回归后，该流程会针对提议的参数值模拟小批次数据，并根据所得的批次残差差异赋予核权重，从而得到一种自归一化伪后验。该伪后验简单、易于并行化，且只需访问拟合得到的回归系数，无需访问原始观测数据。我们将这一构造形式化为总体目标的一个重要性抽样近似；该总体目标对模拟器随机性取平均。我们证明，随着参数抽样次数增加，该方法具有一致性，并确立了从有限样本估计代理回归时的稳定性。随后，我们刻画了批次大小增加、带宽缩小时的渐近集中性，表明伪后验会集中于由所选投影决定的识别集，从而阐明该方法何时能够实现点识别、何时只能实现集合识别。在一个易于处理的非线性模型以及一项使用 DREAMS 模拟套件的宇宙学标定任务中进行的实验，展示了基于回归的投影在计算上的优势，以及低信息量摘要统计量所带来的可识别性局限。

关键词：伪后验、基于模拟的推断、数据投影、无似然推断、广义贝叶斯推断

©2022 Farahi、Rose 和Torrey。许可协议：知识共享署名4.0国际许可协议，
详见 https://creativecommons.org/licenses/by/4.0/。

1 引言

复杂的科学数值仿真器已成为众多科学领域不可或缺的工具，包括宇宙学（Vogelsberger 等，2020年）、气候（Lai 等，2025年）、流行病学（Heesterbeek 等，2015年）和系统生物学（Kitano，2002年）等。这些模拟器将机制性领域知识编码为随机生成模型，通常通过前向过程指定：给定潜在参数，模拟观测数据，而非采用解析上易于处理的似然函数。当似然函数$p(y \mid \theta)$不可用或计算成本高得令人难以承受时，基于显式似然计算或基于梯度的优化的标准贝叶斯方法便不再能直接适用。因此，人们提出了大量无似然或基于模拟的推断（SBI）技术，仅利用前向数据生成模型模拟得到的数据来推断后验分布 $p(\theta \mid y_{\mathrm{obs}})$（例如，Beaumont，2010年；Csill´ery 等，2010年；Blum 等，2013年；Bissiri 等人，2016；Brehmer，2021年；Zammit-Mangion 等，2025）。

这一领域的早期研究起源于近似贝叶斯计算（Tavar´e 等，1997年；Robert，2014年；Sisson 等，2018年）。经典的近似贝叶斯计算方法通过比较模拟数据与观测数据的低维汇总统计量来近似后验分布，并根据 $s(y^{\mathrm{sim}})$ 与$s(y^{\mathrm{obs}})$ 之间的差异接受或加权参数抽样。拒绝近似贝叶斯计算及其变体以基于核函数的接受规则取代显式似然评估，而更先进的方法则引入回归调整、序贯蒙特卡洛或马尔可夫链蒙特卡洛，以提高效率（Marin 等，2012年；Lintusaari 等，2017年）。尽管近似贝叶斯计算方法的渐近性质已得到充分确立（Frazier 等，2018年），且其概念简单，但如何选择有信息量的汇总统计量、差异空间中的维度灾难，以及容差参数所涉及的计算成本与近似质量之间的权衡，都使得近似贝叶斯计算方法难以应用于大规模经验数据。

经典近似贝叶斯计算和许多现代基于模拟的推断方法有一个共同特点：使用精心设计或学习得到的汇总统计量（或低维嵌入），将复杂的模拟器输出压缩成适用于推断的形式（Fearnhead and Prangle，2012年；Jiang 等，2024年）。尽管基于模拟的推断正朝着更灵活、通常由自动化方法学习得到的表示发展，但在汇总统计量与嵌入之间作出选择仍是一项设计决策，会影响推断的质量和稳健性（Forbes 等，2022；Huang 等，2023）。当完整观测 y 维度很高，或存储、传输、共享成本高昂时，研究者通常会将数据压缩为低维摘要统计量，同时保留系统最显著的特征。例如在宇宙学中，大尺度结构巡天通常会被归纳为功率谱、相关函数，或可观测量之间的标度关系（Huterer and Shafer，2017年）。这类摘要统计量可以将计算和存储成本降低数个数量级，也有助于应对隐私、专有信息或数据治理方面的约束。然而，选择不当的摘要统计量可能会丢弃有助于识别底层参数的信息（Forbes 等，2022）。此外，模拟器一旦使用一组固定的摘要统计量完成校准，当引入其他摘要统计量或出现新的下游任务时，所得的后验样本并不总能轻易复用。

基于模拟的推断文献中的另一条研究脉络，聚焦于摊销神经方法。这类方法直接从模拟器输-

出（Greenberg 等，2019；Cranmer 等，2020；Wang 等，2024；Za
mmit-Mangion 等，2025）中学习灵活的条件密度估计器或替代似然。神经后验估计、神经似然估计和神经比率估计等方法，通常通过在大量模拟数据集合上训练深度神经网络，将 $p(\theta \mid y)$ 或 $p(y \mid \theta)$ 近似到一个归一化常数的精度（Cranmer 等，2015；Papamakarios 和 Murray，2016；Lueckmann 等，2019；Hermans 等，2020）。这些方法能够扩展至高维观测，并支持跨多个数据集进行摊销推断，但往往需要大量训练工作、谨慎的正则化，以及大量模拟数据。它们还会通过先验分布、网络架构和损失函数的选择引入额外的建模假设，并可能产生过度自信或校准不佳的后验近似（Hermans 等，2021）。在实际科学应用中，这可能导致计算密集且透明度较低的工作流程，使得可识别性、不确定性量化、后验集中以及对模型设定错误的敏感性更难分析（Bharti 等，2025）。此外，测量噪声、仪器效应、校准误差和选择偏差等实验因素，可能导致经验观测数据与理想化的模拟器输出存在系统性差异。因此，仅使用干净的模拟数据训练神经模拟器型贝叶斯推断方法，在应用于真实数据时可能造成推断偏差或校准不当。

为缓解这些挑战，有人提出使用直接拟合观测数据的代理模型，在模拟器型贝叶斯推断框架中定义损失函数（例如参见 Lilie 等，2025；Rose 等，2025a、2025b；Garcia 等，2025）。这些方法的动机源于这样一种观察：许多科学分析依赖变量之间近似线性的关系，例如天体物理学中的标度关系或流行病学中的剂量-反应曲线。它们不对完整的似然或后验分布进行近似，而是估计伪后验分布。具体而言，这些方法假设一个以 Y 为因变量、以协变量 X为自变量的代理线性回归，其回归系数β° 根据经验数据分布进行估计。随后，从模拟器生成分批模拟数据集，在每个批次内计算平均残差 $Y_{\text{sim}} - \beta^{\circ\top}\mathbf{X}^{\text{sim}}$，并根据该残差均值的大小为每个参数抽样赋予权重。在参数空间上得到的自归一化测度被称为伪后验分布。这与贝叶斯间接似然 （更广义地说，即间接推断）方法相似（Heggland 和 Frigessi，2004；Drovandi 等，2015）：这些方法用一个易于拟合的辅助 模型替代生成模型难以处理的似然，然后通过辅助拟合来衡量观测数据与模拟数据之间的一致性。伪后验分布是一种刻意保持最简的构造。它不试图近似完整的数据似然，也不要求学习嵌入。相反，它将残差均值

$$\mu^\circ(\theta) = \mathbb{E}\big[Y - \beta^{\circ\top}\mathbf{X} \mid \theta\big]$$

作为一种一维差异，用于衡量模拟器对数据观测线性投影的再现程度。 $\mu^\circ(\theta)$接近零的参数会被提高权重；会导致系统性偏差的参数则会被降低权重。从实践角度看，该方法兼具多项颇具吸引力的特性：易于实现，随模拟批次数量线性扩展，可轻松实现并行化，并且只需要观测数据的低维摘要。因此，在模拟器成本高昂且数据共享受限的场景中，该方法尤其具有吸引力，例如，

当经验数据属于专有数据、涉及敏感信息或受严格治理规则约束时。这使得该方法属于无似然方法的一类，但更加注重可解释性、计算简便性和最小化数据访问：一旦已知 β° ，校准时便不再需要原始观测数据集。

尽管这一策略直观、可扩展且易于解释，并且无需共享原始数据，但目前仍缺乏形式保证。具体而言，底层参数在什么条件下可识别、由此得到的方法是否渐近一致，以及替代模型设定错误或有限样本效应会如何影响推断，这些问题仍不清楚。尽管缺乏理论保证，这些方法已逐渐受到关注，近期也已应用于若干实证研究（Busillo等人，2023年；Lilie等人，2025年；Rose 等，2025年a、2025年b；Garcia 等，2025年；Busillo等人，2025年）[1]。鉴于实践中对这类方法的兴趣日益增长，本文对其进行正式研究，并确立所提出的方法是对无似然推断工具箱的一种简单而有原则的补充。我们首先将伪后验的构造形式化为一种分批重要性抽样方案，并将其置于更广泛的基于模拟的推断框架中。接下来，我们对伪后验进行详细的理论分析。在温和的矩条件和连续性假设下，我们证明，基于分批模拟的蒙特卡洛估计量对于定义明确的总体目标是一致的 $\pi_{M,\,\tau}$，并且 $\pi_{M,\,\tau}$ 本身会随着批次大小 M 增大、核带宽 τ以适当速率缩小而收敛至理想化的极限测度。当辅助系数固定为其概率极限 β°时，我们证明总体伪后验会集中于一个识别集

$$\Theta^{\dagger} = \arg\min_{\theta \in \Theta} \mu^{\circ}(\theta)^2,$$

并证明了该集合包含真实数据生成参数$\Theta^{\star}$的条件。这些结果阐明了伪后验所估计的对象、模拟噪声如何通过加权方案传播，以及可识别性如何由所选投影决定。

本文其余部分组织如下。第 2 节介绍形式化的问题设置，定义分批模拟方案，并描述产生伪后验的核加权。第 3 节建立理论基础，包括分批残差的中心极限结果、经验测度的一致性，以及相对于辅助回归估计的稳定性。第 4 节聚焦于固定辅助系数下的可识别性与集中性，刻画识别集，并建立总体伪后验与经验伪后验的集中性。第 6 节基于 DREAMS 数据（Rose 等人，2025年c）给出一项宇宙学应用，突出该方法在现实科学情境中的表现。我们在第 7 节作结，讨论实际意义、局限性，以及将伪后验与更一般的近似贝叶斯计算和基于模拟的推断工作流程相结合的方向。

1. 我们注意到，尽管这些研究采用的思路相似，但计算后验分布/最佳拟合的具体方法有所不同。

2 问题设置

观测数据与增广预测变量。令 $\mathcal{D}_{\text{obs}} = \{(x_i^{\text{obs}}, y_i^{\text{obs}})\}_{i=1}^n$ 表示观测样本，其中 $x_i^{\text{obs}} \in \mathbb{R}^d$ 且 $y_i^{\text{obs}} \in \mathbb{R}$。为编码截距，定义增广预测变量

$$\mathbf{x}_i^{\text{obs}} = (1,\, x_i^{\text{obs}\top})^\top \in \mathbb{R}^{d+1}, \qquad \mathbf{X}_{\text{obs}} = \begin{bmatrix} \mathbf{x}_1^{\text{obs}\top} \\ \vdots \\ \mathbf{x}_n^{\text{obs}\top} \end{bmatrix} \in \mathbb{R}^{n\times(d+1)}, \qquad \mathbf{y}_{\text{obs}} = \begin{bmatrix} y_1^{\text{obs}} \\ \vdots \\ y_n^{\text{obs}} \end{bmatrix} \in \mathbb{R}^n.$$

替代线性模型（对观测数据进行普通最小二乘法回归）。我们拟合代理模型

$$\mathbf{y}_{\text{obs}} = \mathbf{X}_{\text{obs}}\,\beta_{\text{obs}} + \varepsilon, \qquad \mathbb{E}[\varepsilon] = 0,\ \mathbf{Var}(\varepsilon) = \sigma^2 I_n, \tag{1}$$

并定义最小二乘估计量

$$\widehat{\beta} \equiv \widehat{\beta}_n = \arg\min_{\beta\in\mathbb{R}^{d+1}} \|\mathbf{y}_{\text{obs}} - \mathbf{X}_{\text{obs}}\beta\|_2^2 = \left(\mathbf{X}_{\text{obs}}^\top \mathbf{X}_{\text{obs}}\right)^{-1}\mathbf{X}_{\text{obs}}^\top \mathbf{y}_{\text{obs}} \quad \text{（假设可逆）。} \tag{2}$$

令 $\beta^\circ = \text{plim}_{n\to\infty}\,\beta_n$ 表示观测分布 $\mathcal{L}_{\text{obs}}(X, Y)$ 下的最佳线性投影。除非另有说明，第 2–3 节将 β 视为固定；第 3.4 节则允许 $\beta \to \beta^\circ$。例如，在天文学领域，β 可以表示光度—旋转关系的经验观测斜率和截距（Steinmetz 和 Navarro，1999年），也可以表示星系团内的星系速度弥散和热气体温度（Farahi 等人，2018年）。尽管这些回归纯属现象学模型，忽略了数据中任何复杂的非线性或异方差性，但它们提供了一种易于计算且便于解释的低维汇总统计量。

模拟器与分层数据生成。令 $\theta \in \Theta$表示潜在（天文学和物理学背景下的物理）模拟器参数，其先验分布为 $p(\theta)$；对于每个 θ，模拟器生成

$$\begin{aligned} \theta &\sim p(\theta), \\ X^{\text{sim}} \mid \theta &\sim p(x \mid \theta), \\ Y^{\text{sim}} \mid X^{\text{sim}}, \theta &\sim p(y \mid x, \theta), \end{aligned} \tag{3}$$

并扩充$\mathbf{X}^{\text{sim}} = (1,\ X^{\text{sim}\top})^\top \in \mathbb{R}^{d+1}$。在宇宙学背景下，这些参数可能包括物质密度 Ω_m、初始涨落的振幅 σ_8、重子反馈效率，以及控制恒星形成和反馈的其他子网格参数（Weinberg 等人，2013年）。我们的目标是找出那些 θ，使模拟器能够复现由 β概括的观测线性趋势。

批量模拟。令 $n_\theta \geq 1$ 表示用于构建经验测度的独立参数抽样数量。抽取

$$\theta_1, \ldots, \theta_{n_\theta} \overset{\text{iid}}{\sim} p(\theta).$$

对于每个 $j \in \{1, \ldots, n_\theta\}$，生成一个独立批次

$$\{(X_{j,m}^{\text{sim}}, Y_{j,m}^{\text{sim}})\}_{m=1}^M, \qquad X_{j,m}^{\text{sim}} \sim p(x \mid \theta_j), \quad Y_{j,m}^{\text{sim}} \sim p(y \mid x, \theta_j),$$

其中 $M \geq 1$ 表示批次大小，即每个参数值对应的模拟样本对数量。我们设置 $\mathbf{X}_{j,m}^{\text{sim}} = (1,\ X_{j,m}^{\text{sim}\top})^\top$。

^批次差异（残差均值）。给定 β，对每个 θ 定义（总体）残差均值和方差

$$\mu(\theta) := \mathbb{E}\big[Y - \widehat{\beta}^\top \mathbf{X} \mid \theta\big], \qquad v(\theta) := \mathbf{Var}\big(Y - \widehat{\beta}^\top \mathbf{X} \mid \theta\big), \tag{4}$$

其中，期望是在 $p(x \mid \theta)p(y \mid x,\ \theta)$ 下计算的，且 $\mathbf{X} = (1,\ X^\top)^\top$。对于抽样 θ_j，经验批次差异是批次平均残差

$$R_{\theta_j} := \frac{1}{M} \sum_{m=1}^{M} \Big(Y_{j,m}^{\mathrm{sim}} - \widehat{\beta}^\top \mathbf{X}_{j,m}^{\mathrm{sim}}\Big). \tag{5}$$

核加权和伪后验分布。固定带宽 $\tau > 0$，并定义未归一化权重和归一化权重

$$\widetilde{w}_j = \exp\Big(-\tfrac{R_{\theta_j}^2}{2\tau^2}\Big), \qquad w_j = \frac{\widetilde{w}_j}{\sum_{k=1}^{n_\theta} \widetilde{w}_k}. \tag{6}$$

由此得到的自归一化经验测度

$$\Pi_{n_\theta}(\mathrm{d}\theta) = \sum_{j=1}^{n_\theta} w_j\, \delta_{\theta_j}(\mathrm{d}\theta) \tag{7}$$

作为关于 θ 的替代后验分布。由于不存在似然评估，该方法属于无似然或基于模拟的推断技术家族。现在，我们将建立其理论性质。

对于总体分析，引入期望总体权重会很方便

$$L_M(\theta;\tau) := \mathbb{E}\left[\exp\left\{-\frac{R_{M,\theta}^2}{2\tau^2}\right\}\middle|\theta\right], \qquad R_{M,\theta} = \frac{1}{M}\sum_{m=1}^{M}\Big(Y_m^{\mathrm{sim}} - \widehat{\beta}^\top \mathbf{X}_m^{\mathrm{sim}}\Big), \tag{8}$$

以及相应的总体目标

$$\pi_{M,\tau}(\mathrm{d}\theta) \ \propto\ p(\theta)\, L_M(\theta;\tau)\, \mathrm{d}\theta. \tag{9}$$

第 3 节分析了 Π_{n_θ} 作为对 $\pi_{M,\,\tau}$ 的重要性抽样近似，为 R_θ 建立了高斯控制，将自归一化一致性确立为 $n_\theta \to \infty$，证明了 $\beta \to \beta^\circ$ 下的稳定性，并证明了 $M \to \infty$ 在 $\tau \downarrow 0$ 下的集中性。

尽管这种方法与广义贝叶斯方法（Bis- siri 等，2016）有相似之处，且其渐近性质得到了恰当刻画（Miller，2021），但二者在两个关键方面有所不同。首先，这种更新并非通常的吉布斯和广义贝叶斯形式$\pi(\mathrm{d}\theta \mid y_{obs}) \propto p(\theta)\exp\{-\lambda\ell(y_{obs},\ \theta)\}\mathrm{d}\theta$，其中损失是在观测样本上计算的确定性损失。相反，我们的加权方案基于随机批次统计量 $R_{M,\,\theta}$，且总体目标还通过$L_M(\theta;\,\tau)$包含了对模拟器随机性的额外期望。因此，$\pi_{M,\,\tau}$是一个以$(M,\ \tau)$为索引的伪后验，而不是与 y_{obs}的隐式似然相对应的后验分布。其次，观测数据仅通过代入投影 β 参与^构造：该方法并不匹配模拟摘要统计量与

观测摘要统计量，而是通过提高模拟残差均值接近零的参数的权重，来施加近似矩限制。在理想化的大批次情形 $M \to \infty$, $R_{M,\theta}$ 集中于 $\mu(\theta)$附近，而$\pi_{M,\tau}$ 趋近于损失为 $\mu(\theta)^2$ 、温度为 τ 的吉布斯式更新（Martin 和 Syring，2022）；对于有限的 M，它可解释为具有噪声损失的广义贝叶斯更新，或者等价地，解释为将 ABC 式核应用于一维摘要统计量 $R_{M,\theta}$ ，其目标值为 0。这一区别阐明了温度 τ （控制更新强度）与 M （模拟噪声水平）各自的作用，并为第 3 节和第 4 节研究的联合缩放机制提供了依据。

我们还注意到，所提出的方法与近似贝叶斯计算（ABC）的间接似然 表述密切相关（Drovandi 等，2015）。在这类表述中，机理模拟器的似然难以计算，因此会用一种易于拟合的辅助 统计模型来替代。在这类方法中，通过辅助拟合评估观测数据与模拟数据的一致性，从而得到近似（或“伪”）后验分布；该后验分布倾向于支持这样的参数值：在这些参数值下生成的模拟数据能够重现经验分布中存在的辅助结构。具体而言，ABC 间接得分（ABC-IS）及相关的间接推断变体，依据辅助估计方程或得分函数构造差异度量，而非直接使用原始数据层面的距离。这里将基于观测数据拟合得到的回归系数向量 β_{obs} 视为数据集的低维摘要。当在参数值 θ 下生成的模拟数据能够得到与观测数据对应辅助估计值接近的结果时，这些参数值 θ 便会获得更高的后验支持度；通常通过核加权或合成似然构造来实现这一点（Drovandi，2018）。

3 理论结果

本节介绍批量加权方案的理论基础。我们给出关于中心极限定理、一致性、稳定性和集中性的结果。$\widehat{}$除非另有说明，我们将代理估计 β 视为固定；第 3.4 节通过令 β 由一个收敛至其概率极限的估计量估计，放宽这一条件 $\widehat{}$。

3.1 矩结构与中心极限定理

For 为使中心极限定理适用，我们施加以下正则条件 .

假设 3.1（矩正则性、独立性和连续性）固定 $\theta \in \Theta$并定义残差

$$Z = Y - \widehat{\beta}^\top \mathbf{X}, \qquad (X,Y) \sim p(x \mid \theta)p(y \mid x,\theta), \quad \mathbf{X} = (1, X^\top)^\top. \tag{10}$$

对每个 θ:

1. （独立抽样）给定 θ，用于构造 $R_{M,\theta} = \frac{1}{M}\sum_{m=1}^{M} Z_m$ 的模拟残差 $\{Z_m\}_{m=1}^{M}$ 独立同分布，且 —

$$\mu(\theta) := \mathbb{E}[Z \mid \theta], \qquad v(\theta) := \mathbf{Var}(Z \mid \theta) \in [0,\infty). \tag{11}$$

2. ***(***有限矩***)*** 残差具有有限四阶矩，

$$\mathbb{E}\big[|Z|^4 \mid \theta\big] < \infty. \tag{12}$$

3. (连续性) 映射 $\theta \mapsto \mu(\theta)$ 和 $\theta \mapsto v(\theta)$ 在 Θ 上连续。

根据假设 3.1，指数为 2 的 Lyapunov 条件得到满足，因此可对 $R_{M,\,\theta}$ 应用 Lindeberg–Feller 中心极限定理（Billingsley，2017年）。由于 $R_{M,\,\theta}$是 M 个独立同分布残差的均值，Lindeberg–Feller 中心极限定理给出：

引理3.2 令 $s^2(\theta) = v(\theta)/M$。根据假设 *3.1*，

$$R_{M,\theta} \xrightarrow{d} \mathcal{N}\big(\mu(\theta), s^2(\theta)\big) \quad as\ M \to \infty. \tag{13}$$

此外，对于任意有界连续函数 φ，

$$\mathbb{E}[\varphi(R_{M,\theta}) \mid \theta] \longrightarrow \int \varphi(t)\,\mathcal{N}(t \mid \mu(\theta), s^2(\theta))\,\mathrm{d}t. \tag{14}$$

3.2 总体权重函数

回顾总体权重的期望

$$L_M(\theta;\tau) := \mathbb{E}\Big[\exp\big\{-R_{M,\theta}^2/(2\tau^2)\big\} \mid \theta\Big]. \tag{15}$$

根据引理3.2中的高斯近似，当$R_{M,\,\theta}$ 服从正态分布时，我们可以显式计算 L_M。下面的引理使用高斯变量平方的矩母函数。

Lemma 3.3 设 $R_{M,\theta} \sim \mathcal{N}\big(\mu(\theta)\ s^2(\theta)\big)$ 满足 $s^2(\theta) = v(\theta)/M$并固定 $\tau > 0$。n

$$L_M(\theta;\tau) \;=\; \mathbb{E}\left[\exp\left\{-\frac{R_{M,\theta}^2}{2\tau^2}\right\}\middle|\theta\right] \;=\; \sqrt{\frac{\tau^2}{\tau^2 + s^2(\theta)}}\;\exp\left\{-\frac{\mu(\theta)^2}{2\big(\tau^2 + s^2(\theta)\big)}\right\}.$$

大样本M 极限。 随着每个参数对应的模拟抽样次数增加，批次均值 $R_{M,\theta}$ 的抽样方差 $s^2(\theta) = v(\theta)/M$ 趋于零，而引理 3.3 中的高斯近似越来越集中于其均值 $\mu(\theta)$ 附近。因此，

$$L_M(\theta;\tau) = \sqrt{\frac{\tau^2}{\tau^2 + s^2(\theta)}}\;\exp\left\{-\frac{\mu(\theta)^2}{2\big(\tau^2 + s^2(\theta)\big)}\right\} \xrightarrow[M\to\infty]{} \exp\left\{-\frac{\mu(\theta)^2}{2\tau^2}\right\} =: L_\infty(\theta;\tau).$$

直观而言，随着批次平均值中的蒙特卡罗噪声消失，总体权重仅取决于确定性的残差均值 $\mu(\theta)$ 和核带宽 τ。这一极限形式会降低模拟残差均值 $\mu(\theta)$偏离零的参数 θ 的权重，因此量化了模拟器重现由 β概括的观测线性关系的程度。

引理 3.4（权重的一致收敛） 假设 $\sup_{\theta\in\Theta} v(\theta) =: V < \infty$。那么对于任意固定的 $\tau > 0$,

$$\sup_{\theta\in\Theta} \big|L_M(\theta;\tau) - L_\infty(\theta;\tau)\big| \;\longrightarrow\; 0 \qquad as\ M \to \infty.$$

此外，该收敛关于 $\mathcal{O}(1/M)$ θ 是一致的。

3.3 蒙特卡洛一致性

我们将相应的总体伪后验定义为

$$\pi_{M,\tau}(\mathrm{d}\theta) \;\propto\; p(\theta)\, L_M(\theta;\tau)\,\mathrm{d}\theta, \qquad \pi_{\infty,\tau}(\mathrm{d}\theta) \;\propto\; p(\theta)\, L_\infty(\theta;\tau)\,\mathrm{d}\theta. \tag{16}$$

在大规模M情形下，$\pi_{M,\,\tau}$因此近似于理想化测度的平滑版本$\pi_{\infty,\,\tau}$，其围绕 $\mu(\theta)$的集中性 $=0$ 决定了真实参数集的渐近识别。这里，我们证明经验伪后验 Π_{n_θ} 近似于总体目标 $\pi_{M,\,\tau}$。现在我们给出这一重要性抽样估计的大数定律。

定义 3.5 对于一个可测的 $h\colon \Theta \to \mathbb{R}$，满足 $\int |h(\theta)|\,\pi_{M,\,\tau}(\mathrm{d}\theta) < \infty$ 以及 $\int |h(\theta)|\,\pi_{\infty,\,\tau}(\mathrm{d}\theta) < \infty$，定义

$$\Phi_M(h) \;:=\; \int h(\theta)\,\pi_{M,\tau}(\mathrm{d}\theta), \qquad \Phi_\infty(h) \;:=\; \int h(\theta)\,\pi_{\infty,\tau}(\mathrm{d}\theta).$$

定理 *3.6*（蒙特卡洛一致性）设 $h\colon \Theta \to \mathbb{R}$ 可测，且满足 $\int |h(\theta)|\,\pi_{M,\,\tau}(\mathrm{d}\theta) < \infty$。假设对于 $\Theta \sim p(\theta)$，$\mathbb{E}[L_M(\Theta;\,\tau)] > 0$，并且 $\mathbb{E}[L_M(\Theta;\,\tau)\,|h(\Theta)|] < \infty$。令 $\theta_1, \ldots, \theta_{n_\theta} \overset{iid}{\sim} p(\theta)$，并定义自归一化权重 $w_j = \widetilde{w}_j / \sum_{k=1}^{n_\theta} \widetilde{w}_k$，其中 $\widetilde{w}_j = \exp\!\left(-R_{\theta_j}^2/(2\tau^2)\right)$。则

$$\sum_{j=1}^{n_\theta} w_j\, h(\theta_j) \;\xrightarrow[n_\theta\to\infty]{a.s.}\; \int h(\theta)\,\pi_{M,\tau}(\mathrm{d}\theta).$$

所提方法的一个优势是便于重复用于后验分布评估。一旦加权样本 $\{(\theta_j, w_j)\}$ 构建完成，它便作为替代后验发挥作用：任何泛函 $\int h(\theta)\,\pi(\mathrm{d}\theta)$都可由 $\sum_j w_j h(\theta_j)$近似，且任何新的预测观测量都可通过在每个 θ_j下进行前向模拟来评估。引入新的汇总统计量或预测量时，无需重复校准。相比之下，许多无似然方法在引入新的摘要统计量时必须重新运行校准步骤。

命题 3.7 除非 $L_M(\theta;\,\tau) = L_\infty(\theta;\,\tau)$ 对 p-几乎每个 θ 成立（例如，当 $v(\theta) \equiv 0$ 或 $\tau \to \infty$ 时），存在有界可测的 h，使得 $\Phi_M(h) \neq \Phi_\infty(h)$。

证明 若 $L_M \neq L_\infty$ 在一个具有正 p-测度的集合上成立，则它们所诱导的概率测度 $\pi_{M,\tau}$ 与 $\pi_{\infty,\tau}$ 不同。根据全变差的定义，存在可测集 A，使得 $\pi_{M,\tau}(A) \neq \pi_{\infty,\tau}(A)$。取 $h = \mathbb{1}_A$（有界且可测）可得 $\Phi_M(h) \neq \Phi_\infty(h)$。
■

这意味着，一般而言，在有限的 M 下，Φ_M 并不是 Φ_∞ 的无偏代理，但当 $M \to \infty$ 时，Φ_M 会收敛到 Φ_∞。

定理 *3.8* 假设 $\sup_{\theta\in\Theta} v(\theta) < \infty$，固定 $\tau > 0$，并假设 $E[\,|h(\Theta)|\,L_\infty(\Theta;\,\tau)\,] < \infty$ *for* $\Theta \sim p$。那么 $\Phi_M(h) \to \Phi_\infty(h)$ *as* $M \to \infty$。此外，若 $\sup_\theta |L_M(\theta;\,\tau) - L_\infty(\theta;\,\tau)| = \mathcal{O}(M^{-1})$，如引理 *3.4* 所述，则对于有界的 h, $|h| \le K < \infty$，对所有 $\theta \in \Theta$，我们有

$$|\Phi_M(h) - \Phi_\infty(h)| \;=\; \mathcal{O}\!\left(\frac{1}{M}\right).$$

注 3.9 若 $\int |h|\, L_\infty \,\mathrm{d}p < \infty$ 且 $\int |h|\, |L_M - L_\infty| \mathrm{d}p \le C \,\|L_M - L_\infty\|_\infty$，并且 C 有限，则满足 $\mathcal{O}(M^{-1})$ 速率。

通过截断，同样的收敛速率可在任意小的尾部误差范围内成立（参见注 3.9）。

推论 3.10（两阶段收敛） 设 $\widehat{\Phi}_{n_\theta,M}(h) = \sum_{j=1}^{n_\theta} w_j h(\theta_j)$ 为权重如方程 (6)所示的自归一化估计量。在定理 *3.6* 和定理 *3.8* 的可积性假设下，

$$\widehat{\Phi}_{n_\theta,M}(h) \xrightarrow[n_\theta\to\infty]{\text{a.s.}} \Phi_M(h) \quad \textit{for each fixed } M, \qquad \textit{and} \qquad \Phi_M(h) \xrightarrow[M\to\infty]{} \Phi_\infty(h).$$

因此，若 $n_\theta \to \infty$ 且 $M \to \infty$，则 $\Phi_{n_\theta,M}(h) \to \Phi_\infty(h)$。

注 3.11 有限n_θ时，自归一化重要性采样对 $\Phi_M(h)$ 通常存在偏差，但它是强相合的（定理 *3.6*）。同样，正如预期，有限 M 时，$\Phi_M(h)$ 通常不等于 $\Phi_\infty(h)$（命题 *3.7*）。尽管有限(n_θ, M)时，关于 $\pi_{\infty,\tau}$的无偏性并不成立，但定理 *3.6* 和命题 *3.7* 证明了其渐近无偏性。

3.4 稳定性保证

前述结果将 β 视为固定值。实际应用中，β 是通过公式（1）根据观测数据估计得到的。令 β° 表示当观测样本量$n \to \infty$ 趋于无穷时 β 的概率极限（即数据生成过程下，将 Y 对$\mathbf{X}$ 进行最佳线性投影所得的结果）。定义$\mu^\circ(\theta) = \mathbb{E}[Y - \beta^{\circ\top}\mathbf{X} \mid \theta]$, $v^\circ(\theta) = \mathbf{Var}(Y - \beta^{\circ\top}\mathbf{X} \mid \theta)$。

假设 3.12（连续性与包络） 存在 $\mathcal{N}$ 的一个邻域 β°，使得对所有 $\beta \in \mathcal{N}$ 和 $\theta \in \Theta$,

1. $\mu_\beta(\theta) := \mathbb{E}[Y - \beta^\top X \mid \theta]$ 和 $v_\beta(\theta) := \mathit{Var}(Y - \beta^\top X \mid \theta)$ 关于（β，θ）联合连续；*2.* 存在一个可积包络 g，使得对所有 $\beta \in \mathcal{N}$，$L_M^{(\beta)}(\theta;\, \tau) \le g(\theta)$，其中$L_M^{(\beta)}(\theta;\, \tau)$表示将 $\widehat{\beta}$ 替换为 β后的 L_M。

定理 3.13（估计替代参数下的稳定性） 在假设 *3.1* 和 *3.12* 下，若 $\widehat{\beta} \xrightarrow{P} \beta^\circ$，则对于任意有界连续的 $h:\ \Theta \to \mathbb{R}$,

$$\int h(\theta)\, \pi_{M,\tau}^{(\widehat{\beta})}(\mathrm{d}\theta) \xrightarrow{P} \int h(\theta)\, \pi_{M,\tau}^{(\beta^\circ)}(\mathrm{d}\theta),$$

其中 $\pi_{M,\,\tau}^{(\beta)}(\mathrm{d}\theta) \propto p(\theta)\, L_M^{(\beta)}(\theta;\, \tau)\, \mathrm{d}\theta$ 且 $L_M^{(\beta)}(\theta;\, \tau) := \mathbb{E}\left[\exp\left\{-\frac{R_M^{(\beta)\,2}}{2\tau^2}\right\} \middle| \theta\right]$ 其中 $R_{M,\,\theta}^{(\beta)} = M^{-1}\sum_{m=1}^{M}\left(Y_m - \beta^\top \mathbf{X}_m\right)$。

定理 3.13 证明，伪后验的极限行为对替代参数的估计误差保持稳定。β。尽管 β 是通过估计得到的

有限数据，只要它依概率收敛到其概率极限 β°，且矩函数关于 $(\beta,\ \theta)$ 连续变化，θ 上的后验测度就会依概率收敛到与已知真实替代参数 β° 时相同的极限。该结果保证，从观测数据估计 β 所带来的不确定性不会渐近地传递到关于 θ的推断层：Θ上的后验质量渐近地保持不变，不受替代参数拟合中的微小扰动影响。

结合定理 3.6 和 3.13 可知，当 $n_\theta \to \infty$ 和观测样本量 $n \to \infty$ 均满足相应条件时，它们会联合收敛。这保证经验伪后验 Π_{n_θ} 能够逼近理想化目标 $\pi^{(\beta^\circ)}{}_{M,\tau}$。代理回归的一致性足以确保整体推断程序的稳定性，从而为伪后验方法所依据的两阶段估计框架提供了依据。

4 可识别性与集中性质

迄今为止，我们基于模拟批次和核加权构建了伪后验框架，证明了在替代参数固定时该框架的收敛性，并证明当代理拟合 β° 被其估计值 β 替代时，$\widehat{\ }$ 方法仍保持稳定。现在，我们将重点考察替代系数固定为该极限值时所得测度的总体性质。在此设定下，通过最佳线性投影 β° 所对应的残差均值和方差来评估模拟器参数 θ，从而得到量 $\mu^\circ(\theta)$ 和 $v^\circ(\theta)$。本节的主要目标是证明，随着模拟抽样次数和批次大小增加，伪后验会集中。这些结果为替代后验 Π_{n_θ} 逼近其总体对应量 $\pi_{M,\tau}$ 提供了正式的统计学依据；同时，当模拟噪声和 β 不确定性消失时，$\pi_{M,\tau}$ 本身会集中于识别集 $\Theta^\dagger$ 周围。我们将证明，在可识别性假设下，该方法在大样本极限下能够恢复正确的识别参数区域。

在本节中，代理系数固定为其概率极限 β°；所有依赖于 β 的对象均在 β° 处求值，并在不致混淆时省略上标。回忆

$$\mu^\circ(\theta) = \mathbb{E}\big[Y - \beta^{\circ\top}\mathbf{X} \mid \theta\big], \qquad v^\circ(\theta) = \mathbf{Var}\big(Y - \beta^{\circ\top}\mathbf{X} \mid \theta\big), \tag{17}$$

对于批次大小 M 和带宽 $\tau > 0$,

$$L_M(\theta;\tau) \ := \ \mathbb{E}\left[\exp\left\{-\frac{R_{M,\theta}^2}{2\tau^2}\right\}\middle|\theta\right], \quad R_{M,\theta} = \frac{1}{M}\sum_{m=1}^{M}\big(Y_m^{\mathrm{sim}} - \beta^{\circ\top}\mathbf{X}_m^{\mathrm{sim}}\big). \tag{18}$$

总体目标为

$$\pi_{M,\tau}(\mathrm{d}\theta) \ \propto \ p(\theta)L_M(\theta;\tau)\,\mathrm{d}\theta, \tag{19}$$

以及基于独立同分布抽样 $\theta_1,\dots,\theta_{n_\theta} \sim p$ 的自归一化蒙特卡罗近似,

$$\Pi_{n_\theta}(\mathrm{d}\theta) \ = \ \sum_{j=1}^{n_\theta} w_j\,\delta_{\theta_j}(\mathrm{d}\theta), \qquad w_j = \frac{\tilde{w}_j}{\sum_{k=1}^{n_\theta}\tilde{w}_k}, \quad \tilde{w}_j = \exp\left\{-\frac{R_{\theta_j}^2}{2\tau^2}\right\}. \tag{20}$$

令识别集为

$$\Theta^{\dagger} = \arg\min_{\theta\in\Theta} \mu^{\circ}(\theta)^2$$

定义为使总体差异$\mu^{\circ}(\theta)^2$达到最小值的参数值集合。在紧集 Θ 上， $\mu^{\circ}(\theta)$的连续性保证最小值存在，但不保证其唯一性。因此， $\Theta^{\dagger}$ 可能是一个维数大于零的低维流形。在这种情况下，基础模拟器和代理模型相对于观测到的矩条件只能实现部分识别，因此得到的是集合识别 而非点识别。后续分析假设识别集 $\Theta^{\dagger}$ 是一个紧集，并以概率质量向该识别集收敛的方式表述所有渐近集中结果。

假设 4.1（一致矩正则性） Θ是紧集，并且存在常数$C < \infty$、 $\delta > 0$和 $0 < v_{\min} \le v_{\max} < \infty$，使得对所有 $\theta \in \Theta$,

$$\sup_{\theta\in\Theta} \mathbb{E}\Big[|Z|^{2+\delta} \mid \theta\Big] \le C, \quad v^{\circ}(\theta) \in [v_{\min}, v_{\max}], \tag{21}$$

其中 $Z = Y - \beta^{\circ\top}\mathbf{X}$和$(X、 Y) \sim p(x \mid \theta)p(y \mid x、 \theta)$。此外， $\mu^{\circ}(\theta)$和 $v^{\circ}(\theta)$在 Θ上连续。

假设 4.2（精确可表示性）存在 $\theta^{\star} \in \Theta$ 使得 $\mathcal{L}_{\text{sim}}(X, Y \mid \theta^{\star}) = \mathcal{L}_{\text{obs}}(X, Y)$，其中 $\mathcal{L}_{\text{obs}}(X, Y)$和 $\mathcal{L}_{\text{sim}}(X, Y \mid \theta)$分别是观测分布和数据生成分布。

假设 4.2 意味着数据生成模型被正确指定。迄今为止，我们尚未作出这一假设，但集中性保证需要作出这一假设。

假设 4.3（可识别性）令 $\Theta^{\dagger} = \text{argmin}_{\theta\in\Theta}\, \mu^{\circ}(\theta)^{2},$，它非空且紧致。令先验分布 $p(\theta)$在 $\Theta^{\dagger}$ 的某个邻域上连续且严格为正，并且对每个开集 $U \supset \Theta^{\dagger}$，都存在一个间隙 $\eta(U) > 0$，使得

$$\inf_{\theta\notin U} \mu^{\circ}(\theta)^2 \ge \inf_{\vartheta\in\Theta} \mu^{\circ}(\vartheta)^2 + \eta(U). \tag{22}$$

引理 4.4在假设 *4.2* 下，我们有

$$\inf_{\vartheta\in\Theta} \mu^{\circ}(\vartheta)^2 = 0 \qquad \textit{and} \qquad \theta^{\star} \in \Theta^{\dagger}.$$

这是由普通最小二乘估计的 $\beta^{\circ\top}$ 和$\mathbb{E}_{\text{obs}}\big[Y - \beta^{\circ\top}\mathbf{X}\big] = 0$这一事实直接推出的。尽管逆命题未必成立， $\mu^{\circ}(\theta^{\triangle})^{2} = 0$并不意味着 $\mathcal{L}_{\text{sim}}(X, Y \mid \theta^{\triangle}) = \mathcal{L}_{\text{obs}}(X, Y)$。

引理 4.5 在假设 *4.1* 下，并且令

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, \mathrm{d}u, \tag{23}$$

then

$$\sup_{\theta\in\Theta}\ \sup_{t\in\mathbb{R}}\left|\Pr\left(\frac{R_{M,\theta}-\mu^{\circ}(\theta)}{\sqrt{v^{\circ}(\theta)/M}}\le t\ \middle|\ \theta\right)-\Phi(t)\right|\ \le\ \frac{C'}{\sqrt{M}}, \tag{24}$$

其中常数 C' 仅取决于一致矩界。

引理 4.6 在假设 *4.1* 下，定义有效方差

$$\tau^2_{\mathrm{eff}}(\theta;M,\tau)\ =\ \tau^2+\frac{v^{\circ}(\theta)}{M}.$$

于是，对 $\theta\in\Theta$ 一致地，

$$L_M(\theta;\tau)\ =\ \sqrt{\frac{\tau^2}{\tau^2_{\mathrm{eff}}(\theta;M,\tau)}}\,\exp\!\left\{-\frac{\mu^{\circ}(\theta)^2}{2\,\tau^2_{\mathrm{eff}}(\theta;M,\tau)}\right\}+\ r_{M,\tau}(\theta), \tag{25}$$

其中 上确界$_{\theta\in\Theta}|r_{M,\ \tau}(\theta)|\le C''/\sqrt{M}$，常数 $C''<\infty$ 仅依赖于假设 *4.1* 中的一致矩界。

证明概要。我们写作 $R_{M,\ \theta}=\mu^{\circ}(\theta)+\sqrt{v^{\circ}(\theta)/M}\,Z$，其中 Z 已标准化。对于 Z 的高斯代理，通过配方计算$\mathbb{E}[\exp\{-R_\theta^2/(2\tau^2)\}]$，并利用引理 4.5 一致地控制近似误差；连续性和紧致性使常数具有一致性。

引理 4.7（总体拉普拉斯界）设 $U\supset\Theta^{\dagger}$ 为开集，并令 $\eta=\eta(U)\ >0$ 为假设 *4.3* 中的分离常数。在假设 *4.1* 和 *4.2* 下，存在常数 $C(U)\in(0,\infty)$ 和 $C'''\in(0,\infty)$，它们与 M、τ 无关；并存在常数 $M_0\in\mathbb{N}$ 和 $K>0$，使得对所有 $M\ge M_0$ 及所有满足 $M\tau^2\ge K$ 的 $\tau>0$，

$$\frac{\int_{\Theta\setminus U}p(\theta)\,L_M(\theta;\tau)\,\mathrm{d}\theta}{\int_U p(\theta)\,L_M(\theta;\tau)\,\mathrm{d}\theta}\ \le\ C(U)\,\exp\!\left\{-\frac{\eta}{4\left(\tau^2+v_{\max}/M\right)}\right\}\left(1+\frac{C'''}{\sqrt{M}}\right),$$

备注。温和的耦合 $M\tau^2\ge K$可防止分母中的前置因子 $\sqrt{\tau^2/(\tau^2+v_{\max}/M)}$ 在固定 M 时随 $\tau\downarrow 0$ 收缩。这是核拉普拉斯分析中的标准做法，并且与集中区间 $\tau\downarrow 0,\,M\to\infty$（其中 $M\tau^2\to\infty$）相容。

最后，我们证明主要的集中结果。

定理 4.8（固定 β° 时的集中性） 假设 *4.1–4.3*。令 $M\to\infty$ 和 $\tau=\tau_M\downarrow 0$ 使得

$$M\tau^2\longrightarrow\ \infty. \tag{26}$$

那么，对于任意开 $U\supset\Theta^{\dagger}$ 和任意闭 F，若满足 $F\cap\Theta^{\dagger}=\emptyset$，

$$\pi_{M,\tau_M}(U)\ \longrightarrow\ 1,\qquad \pi_{M,\tau_M}(F)\ \longrightarrow\ 0. \tag{27}$$

此外，如果还满足 $n_\theta\to\infty$，则几乎必然

$$\Pi_{n_\theta}(U)\ \longrightarrow\ 1,\qquad \Pi_{n_\theta}(F)\ \longrightarrow\ 0. \tag{28}$$

备注。 (i) 仅取 $M \to \infty$ 是不够的：如果 $\tau_M \equiv \tau_0 > 0$，那么 $\pi_{M,\ \tau_0}$ 是弥散的，不会集中。(ii) 一个实际可行的充分缩放是 $M\tau_M^2 \to \infty$，它意味着 $\tau_M^2 + v_{\max}/M \to 0$，并使高斯代理保持一致。(iii) 缩放 $M\tau_M^2 \to \infty$ 保证引理 4.7 中的分母下界不会随着 $\tau_M \downarrow 0$ 在有限 M 下退化，并且与所需的状态 $M \to \infty$、$\tau_M \downarrow 0$ 兼容。

回顾一下，β° 表示在观测到的概率律 $\mathcal{L}_{\text{obs}}(X,\ Y)$ 下的最佳线性投影，并且

$$\mu^\circ(\theta) \ := \ \mathbb{E}_{\text{sim}}\big[Y - \beta^{\circ\top}\mathbf{X} \,\big|\, \theta\big], \qquad \Theta^\dagger \ := \ \arg\min_{\theta\in\Theta} \ \mu^\circ(\theta)^2. \tag{29}$$

Let

$$\Theta^\star \ := \ \{\theta \in \Theta : \ \mathcal{L}_{\text{sim}}(X, Y \mid \theta) = \mathcal{L}_{\text{obs}}(X, Y)\} \tag{30}$$

它是这样一个（可能是集合值的）参数集合：其中的模拟器能够复现观测到的联合概率律。这意味着模拟器对数据的设定是正确的，但线性投影仍可能设定错误。

命题 4.9 如果 $\theta \in \Theta^\star$ 蕴含 $\mathcal{L}_{\text{sim}}(X, Y \mid \theta) = \mathcal{L}_{\text{obs}}(X, Y)$，那么

$$\Theta^\star \ \subseteq \ \Theta^\dagger. \tag{31}$$

当残差矩不具分离性时，集合 $\Theta^\dagger$ 可能严格大于 $\Theta^\star$；也就是说，存在满足 $\mu^\circ(\theta) = 0$ 的 $\theta \notin \Theta^\star$（线性摘要导致欠识别）。在这种情况下，伪后验会集中于更大的伪真值集合 $\Theta^\dagger$，它可能是一个低维流形。这一结果揭示了所提出损失函数的关键局限：在大多数实际应用中，点识别并不可行。我们将在接下来的实验中展示这一性质。

5 数值模拟推断实验

我们考虑一个受控的二维基于模拟的推断问题，其中数据生成过程对潜在参数和观测协变量均呈非线性，而用于构造伪后验分布的代理模型则被有意设定为错误模型。本实验旨在比较所提出的批次残差加权方法与精确贝叶斯后验，在这一似然已知且可通过数值方法求值的情境下进行比较。

生成模型。 令 $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$ 表示未知的参数向量。我们设定各向同性高斯先验

$$\pi(\theta) = \mathcal{N}(0, 5^2 I_2). \tag{32}$$

给定 θ，观测数据按照以下层次模型生成。首先，从对数正态分布中抽取潜在协变量 X_i，

$$\log X_i \mid \theta \sim \mathcal{N}\left(\frac{\theta_1}{5},\, 0.5^2\right), \tag{33}$$

这保证了 $X_i > 0$ 几乎必然成立。给定 X_i 和 θ，响应变量 Y_i 的条件均值为

$$\bar{Y}_i(\theta, X_i) = \theta_1 \log X_i + \theta_0 X_i, \tag{34}$$

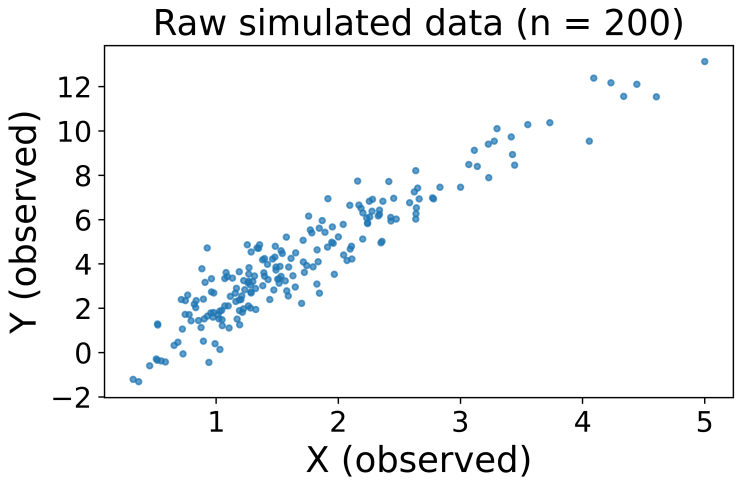


图1：模拟观测数据的一次实现。

并且观测值中含有加性高斯噪声，

$$Y_i \mid X_i, \theta \sim \mathcal{N}\big(\bar{Y}_i(\theta, X_i),\, 1\big)\,, \tag{35}$$

对于 $i = 1,\ldots,n$。在真实参数值 $\theta_{\text{true}} = (2, 2)$ 处模拟观测数据，样本量为 $n_{\text{obs}} = 200$。图5展示了 (X, Y) 的散点图，呈现出明显的曲线趋势，与线性条件均值不一致。

通过 Metropolis–Hastings 计算精确后验分布。由于方程（33）–（35）所隐含的似然在解析上可处理，我们使用随机游走 Metropolis–Hastings（RWMH）采样器计算一个参考后验分布，并以此作为基准来评估我们的伪后验。对于观测数据 $\{(x_i, y_i)\}_{i=1}^{n}$，对数似然为

$$\log p(y \mid x, \theta) = \sum_{i=1}^{n} \log \mathcal{N}(y_i \mid \theta_1 \log x_i + \theta_0 x_i,\, 1)\,, \tag{36}$$

将其与方程（32）的高斯先验结合即可得到后验分布。RWMH 链从先验均值开始初始化，运行 4×10^4 次迭代，并将最初的 5×10^3 个样本作为预热期样本丢弃。其余样本视为从精确后验分布中抽取的样本，并用于与基于模拟的推断伪后验进行定量和定性比较。

替代回归模型。为了定义低维差异，我们引入一个相对于真实生成过程故意设定错误的线性替代模型。具体而言，我们对 Y 关于截距和 X进行普通最小二乘回归，

$$Y \approx \beta_0 + \beta_1 X, \tag{37}$$

$\hat{}\,\hat{}\,\hat{}$ 使用观测数据。估计系数 $\beta = (\beta_0,\ \beta_1)$ 只计算一次，并在整个推断过程中保持固定。该代理模型忽略了方程（34）中对 X 的对数依赖，因此无法精确表示真实的条件均值。

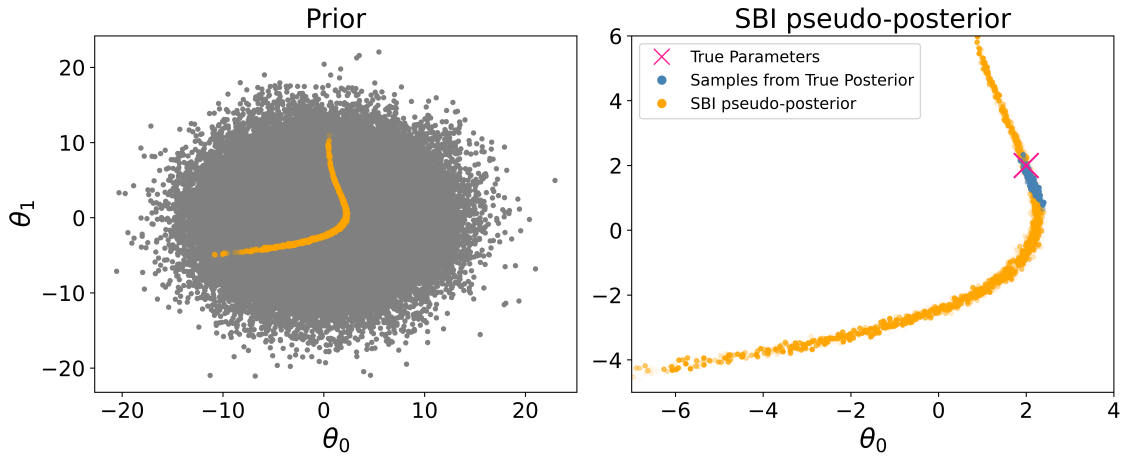


图 2：**左图：** (θ_0, θ_1) 空间中的先验分布样本（灰色），以及由批量残差加权诱导的 SBI 伪后验分布的支撑集（橙色）。先验分布弥散且各向同性，而伪后验分布则集中在一条弯曲的一维流形（识别集）上。**右图：** 通过 Metropolis–Hastings 获得的精确后验分布（蓝色）、SBI 伪后验分布（橙色）与真实参数值 θ_{true} （粉色叉号）的比较。伪后验分布在一个维度上明显集中于真实后验质量附近，而在另一个维度上则无法实现识别。

评估。 我们通过比较 SBI 伪后验与精确后验来评估所提出方法的性能。图 2 展示了分批残差核加权对参数空间的影响。左图显示了从高斯先验中抽取的样本，以及在 SBI 伪后验下获得非可忽略权重的参数子集。先验分布虽然弥散且近乎各向同性，但伪后验质量却收缩到 $(\theta_0,\ \theta_1)$ 空间中的一条弯曲的一维流形上。这种几何形态反映了这样一组参数：它们所产生的模拟数据，其代理回归残差与真实数据中的残差相匹配。

右侧面板将基于模拟的推断伪后验与通过 Metropolis–Hastings 抽样得到的精确贝叶斯后验进行了比较。由于条件均值$\theta_1 \log X + \theta_0 X$存在非线性依赖，$\theta_0$ 与 θ_1的真实后验之间存在很强的相关性。尽管伪后验是使用设定错误的线性代理模型构建的，它仍覆盖了真实后验。然而，伪后验会集中于一条狭窄的一维流形，这表明批量残差准则识别的是那些能够复现观测到的代理模型行为的参数，而非完整似然。该实验表明，伪后验不必处处近似完整的后验密度，而是会集中于一组更低维的参数集合；在选定的代理统计量下，这些参数在观测上无法区分。从这个意义上说，伪后验的几何形态直接揭示了代理模型所导致的可识别性问题与退化现象。这是所提方法的关键局限。

因此，鉴于其计算上的可处理性，该方法适用于探索性分析；但如果目标是估计真实后验，该方法可能会导致可识别性问题。

6 示例：利用星系属性进行宇宙学推断

星系形成的前向模型刻画了天体物理过程与可观测星系属性之间复杂的非线性耦合，因此无法在其高维参数空间中直接进行基于似然的推断（Tortorelli et al., 2025）。DREAMS（暗物质与机器学习天体物理模拟项目；Rose et al., 2025c,b）等模拟套件提供了一个可控环境，可系统地改变不确定的物理参数，从而能够根据观测到的星系标度关系进行统计校准。DREAMS 套件包含 1,024 个采用 IllustrisTNG（TNG；Pillepich et al., 2018；Weinberger et al., 2018）星系形成模型的流体动力学模拟，该模型涵盖了恒星形成、恒星反馈、活动星系核（AGN）反馈、超新星（SN）反馈、化学丰度增丰等多种物理过程。本节利用 DREAMS 中的冷暗物质银河系质量星系模拟套件，找出能够产生物理上合理的星系群体的参数空间区域，并量化当前数据对亚网格物理各方面的约束强度。我们进行了此处介绍的伪后验分析，以确定最能复现观测星系标度关系的参数空间区域，包括恒星质量—晕质量关系（SMHM）和黑洞质量—恒星速度弥散度（M–σ）关系。我们在不同时间测量 SMHM 关系，并采用两组测量来约束这些参数：一组对应当前时期（$z = 0$），另一组对应 ~ 7.5 亿年前（$z = 1$）。SMHM 关系的数据取自 Behroozi 等（2019），而 M–σ 关系的数据取自 Greene 等（2020）。

数据。 用于创建 DREAMS 的 TNG 模型包含 $\mathcal{O}$（100）个会影响星系形成的物理和数值参数。DREAMS 模拟具体改变了三个不确定的天体物理参数和两个会影响星系形成的宇宙学参数。具体而言，这些参数包括物质密度 $\Omega_m \in [0.274, 0.354]$、大尺度物质涨落的振幅 $\sigma_8 \in [0.780, 0.888]$、超新星反馈比能 $\bar{e}_w \in [0.9, 14.4]$、超新星风的速度 $\kappa_w \in [3.7, 14.8]$，以及活动星系核反馈与周围气体的耦合强度 $\epsilon_{f,\,\mathrm{high}} \in [0.025, 0.4]$。因此，$\theta = \{\Omega_m,\ \sigma_8,\ \bar{e}_w,\ \kappa_w,\ \epsilon_{f,\,\mathrm{high}}\}$。宇宙学参数 Ω_m 和 σ_8 对当今星系属性的影响可以忽略，因此我们将分析重点放在约束这三个天体物理参数上。

我们使用 Nguyen 等人（2024年）中介绍的条件归一化流模拟器作为数据生成器，对本文研究的星系属性进行抽样（另见 Rose 等人，2025b）。我们遵循此处介绍的方案，并通过高斯似然函数计算未归一化权重

$$w_i = \exp\left(-\frac{\left(\sum_{i=1}^{M} y_{i,k} - \mu_{\mathrm{obs},i,k}\right)^2}{2M^2\tau_k^2}\right) \tag{38}$$

其中，$y_{i,k}$ 是星系 i 在关系 k中的预测属性，$\mu_{\mathrm{obs},k}$ 是观测均值，而 τ_k 是一个自由温度参数，其设定值为根据观测关系得出的不确定性。

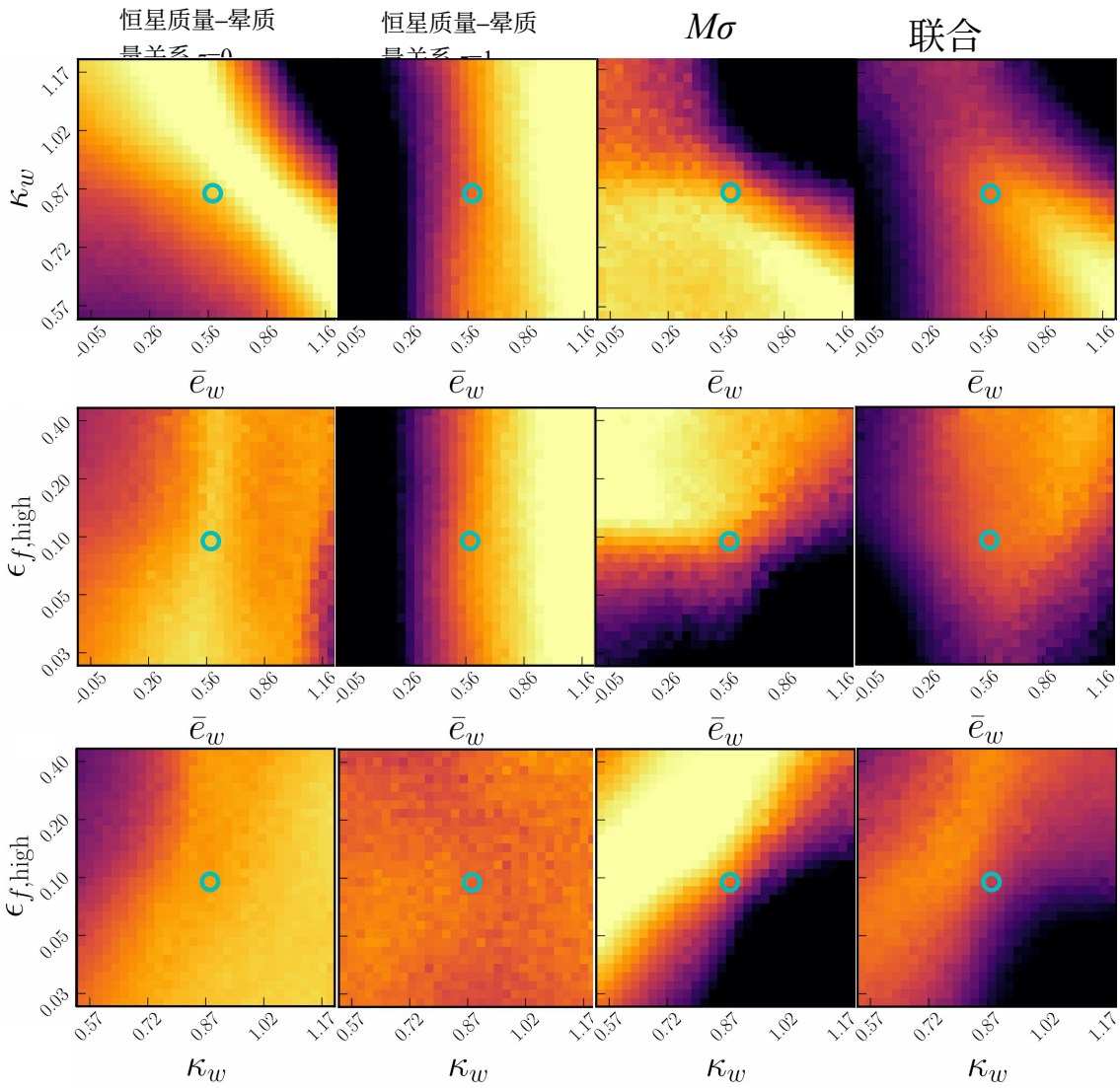


图3：三个不确定天体物理参数的边缘化二维伪后验分布：超新星反馈的比能（$\bar{e}_w$）、超新星风的速度（κ_w）以及活动星系核反馈耦合强度（$\epsilon_{f,\ \mathrm{high}}$）。从左至右，各列分别展示了基于$z=0$处的恒星质量—晕质量关系、 $z=1$处的恒星质量—晕质量关系、黑洞质量–恒星速度弥散度（M–σ）关系以及联合约束得出的约束。颜色越亮表示权重越高的区域，青色环带表示该模型中使用的基准参数。在 $z=0$ 恒星质量—晕质量关系中， $\bar{e}_w$ 与 κ_w 之间的退化会因纳入 $z=1$ 数据而消除。

结果。 图3展示了根据这些权重得出的边缘化伪后验。每一行展示了三个天体物理参数中两个参数的边缘化伪后验——

数， $\bar{e}_w$、 κ_w和$\epsilon_{f,\ \mathrm{high}}$。每列显示来自一种不同标度关系的边缘化约束，最右列则显示所有单独标度关系的综合权重。

在当前时期， $z = 0$、恒星质量—晕质量关系以及 M–σ 关系中的两个超新星反馈参数 $\bar{e}_w$ 和 κ_w，在参数空间的很大一部分区域内存在退化。这种退化是意料之中的，因为这两个参数都会调节气体从恒星形成区的喷出。然而，由于这些星系只有在接近当前时期时才会超过最低风速阈值，风速参数 κ_w在更早时期的影响会减弱。这个速度阈值使早期星系属性几乎不受影响，因此风能参数 $\bar{e}_w$主导了星系属性的变化，这一点可以从 $z = 1$处的恒星质量—晕质量关系中看出。加入来自宇宙早期的这些约束后， $\bar{e}_w$ 和 κ_w之间的退化得以打破；图3右上面板显示出对 κ_w低值和 $\bar{e}_w$ 高值的明显偏好。

由于这些星系质量较低，恒星质量—晕质量关系在任何时期都无法对活动星系核参数 $\epsilon_{f,\mathrm{high}}$施加很强的约束。另一方面， M–σ 关系直接将产生活动星系核反馈的黑洞质量与观测数据进行比较，因此能对 $\epsilon_{f,\mathrm{high}}$施加严格得多的约束，并倾向于较高的取值。这使综合伪后验分布偏好较高的 $\epsilon_{f,\mathrm{high}}$ 和 $\bar{e}_w$ 取值，但 $\epsilon_{f,\mathrm{high}}$ 与 κ_w 之间仍存在退化，需要借助其他标度关系才能消除。结果表明，通过这里的伪后验分布，我们只需付出建模真实数据所需工作量的一小部分，就能获得有意义的物理洞见。

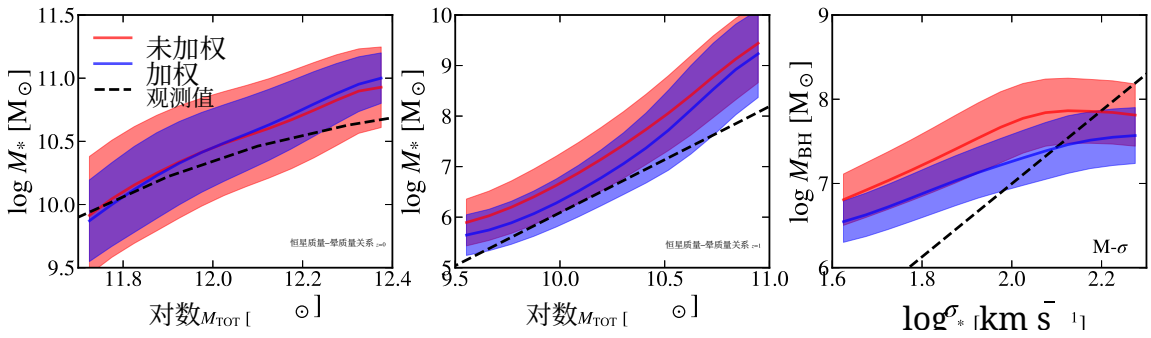


图4：在应用伪后验权重前后，模拟星系标度关系与观测数据的比较。各面板分别展示 $z = 0$ 恒星质量—晕质量关系（左）、 $z = 1$ 恒星质量—晕质量关系（中）以及 M–σ 关系（右）。红色阴影区域表示模拟的未加权分布，蓝色阴影区域表示应用图3最右侧面板所示综合权重后的分布。黑色虚线表示取自Behroozi 等（2019）和Greene 等（2020）的平均观测关系。加权分布，尤其是 $z = 1$ 恒星质量—晕质量关系和M–σ 关系，显著向观测数据偏移，表明伪后验方法能够正确识别与观测关系更为吻合的参数空间区域。

为验证伪后验分布，我们将加权和未加权的星系群体与观测数据进行比较。图4左图展示了 $z=0$ 恒星质量—晕质量关系，中图展示了 $z=1$ 恒星质量—晕质量关系，右图展示了 $M-\sigma$ 关系。加权关系采用图3右栏中的综合权重。对于图4所示的全部三种关系，未加权分布（红色）均偏离观测到的平均关系，这表明，与用于生成该数据集的均匀先验相比，需要对模拟参数施加更严格的约束。

应用权重后，加权的 $z=1$ 恒星质量–晕质量关系和 M–σ 分布（蓝色）向观测平均关系（黑色）移动。应用权重后， $z=0$ 恒星质量–晕质量关系没有显示出显著变化，但仍与观测关系更加一致。这种一致性表明，参数空间中权重较高的区域对应于物理上更为合理的模拟参数空间实现。然而，模拟参数与偏移量之间仍然存在的退化，以及加权关系与观测关系之间仍然存在的偏差表明，还可以加入其他标度关系，以进一步约束这些参数；同时，也可以约束更多模拟参数，使其更好地匹配观测结果。

7 讨论与结论

我们对一种用于基于模拟的推断的基于替代模型的方法进行了理论分析，并展示了其在天体物理学数据分析中的应用。我们的分析确立了中心极限定理、经验伪后验的一致性、关于替代模型回归系数估计的稳定性，以及对能够复现观测线性关系的模拟器参数的集中性。我们的理论结果刻画了所提出方法有效以及可能失效的条件。除具体应用之外，该框架还凸显了简单且易于解释的数据投影汇总统计量——在此即观测定律下的最佳线性投影——如何能够在观测数据与大规模科学模拟器之间提供一种计算高效且隐私保护的接口。由于只需访问替代模型回归系数 $\widehat{\ }\beta$，即使底层观测数据集无法直接共享，该方法仍能实现推断；其原因可能是隐私或治理约束、计算上的不可处理性，或难以准确建模测量误差、样本选择和数据依赖等观测效应。

与此同时，我们的分析凸显了所提出的损失函数的主要局限，并阐明了它在更广泛的无似然推断领域中的恰当定位。由于伪后验是基于单个汇总统计量构建的，因此通常无法保证对模拟器参数进行完全的点识别。我们记录了一些情形：不同的参数值会产生相同的期望残差均值，导致伪后验集中在一个低维流形上，而该流形可能严格大于真实参数集。这体现了基于矩或基于汇总统计量的推断的一项基本特性：可识别性受所选汇总统计量和损失函数所含信息的限制；若要获得更精确的推断，则需要加入额外结构。因此，伪后验最适合被视为一种快速、稳定且透明的校准机制，用于使模拟器与所选数据投影相匹配。它可作为多阶段推断流程中的实用第一层，并与更丰富的汇总统计量相辅相成

、额外的代理模型或表达能力更强的基于模拟的推断技术结合使用，以满足科学目标对更高保真度或完整贝叶斯后验精度的要求。

虽然伪后验计算效率高且能保护隐私，可作为完全基于似然的推断方法的替代方案，也可用于高效的探索性分析、模型检验和初步校准，但不应将其解读为具有与精确贝叶斯后验相同的推断分辨率。相反，它们提供了一个有原则的起点，可与信息量更大的摘要统计量或更先进的基于模拟的推断方法结合，从而在复杂的生成建模任务中获得更高的统计功效和更精确的识别。

致谢

我们感谢 Peter M¨uller、Pratik Patil 和 Abhra Sarkar 提供的宝贵讨论。AF 和 PT 感谢美国国家科学基金会通过合作协议 2421782，以及西蒙斯基金会授予美国国家科学基金会—西蒙斯宇宙起源人工智能研究所（CosmicAI，https://www.cosmicai.org/）的 MPS-AI-00010515 号基金提供支持。

附录 A. 证明

证明 [引理 3 的证明。2] 固定 $\theta \in \Theta$ 并定义

$$Z_m = Y_m^{\mathrm{sim}} - \widehat{\beta}^\top \mathbf{X}_m^{\mathrm{sim}}, \qquad m = 1, \ldots, M.$$

根据假设 3.1(i)，给定 θ 时，$(Z_m)_{m=1}^M$ 为独立同分布，且

$$\mathbb{E}[Z_m \mid \theta] = \mu(\theta), \qquad \mathbf{Var}(Z_m \mid \theta) = v(\theta) < \infty.$$

定义 $R_{M,\theta} = M^{-1} \sum_{m=1}^M Z_m$，并注意到

$$\sqrt{M}\, \frac{R_{M,\theta} - \mu(\theta)}{\sqrt{v(\theta)}} = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \frac{Z_m - \mu(\theta)}{\sqrt{v(\theta)}}.$$

根据假设 3.1(ii)，$\mathbb{E}[|Z_m - \mu(\theta)|^4 \mid \theta] < \infty$。指数为$\delta = 2$ 时的 Lyapunov 比值为

$$\frac{\sum_{m=1}^M \mathbb{E}[|Z_m - \mu(\theta)|^{2+\delta} \mid \theta]}{\Big(\sum_{m=1}^M \mathbf{Var}(Z_m \mid \theta)\Big)^{1+\delta/2}} = \frac{M\, \mathbb{E}[|Z - \mu(\theta)|^4 \mid \theta]}{\big(M\, v(\theta)\big)^2} = \frac{\mathbb{E}[|Z - \mu(\theta)|^4 \mid \theta]}{M\, v(\theta)^2} \xrightarrow[M\to\infty]{} 0.$$

因此 Lyapunov 条件成立，Lindeberg–Feller 中心极限定理适用：

$$\sqrt{M}\, \frac{R_{M,\theta} - \mu(\theta)}{\sqrt{v(\theta)}} \xRightarrow[M\to\infty]{} \mathcal{N}(0,1) \quad (\text{conditional on } \theta).$$

根据变换的线性性，

$$R_{M,\theta} \xRightarrow[M\to\infty]{} \mathcal{N}\big(\mu(\theta),\, v(\theta)/M\big) = \mathcal{N}\big(\mu(\theta),\, s^2(\theta)\big),$$

这证明了第一个结论。设 φ 有界且连续。分布收敛和 Portmanteau定理意味着

$$\mathbb{E}[\varphi(R_{M,\theta}) \mid \theta] \longrightarrow \int \varphi(t)\, \mathcal{N}\big(t \mid \mu(\theta), s^2(\theta)\big)\, \mathrm{d}t,$$

从而确立第二个结论。证明至此完成。 ■

证明 [引理 3.3 的证明。] 假设 $Z := R_{M,\theta} \sim \mathcal{N}(\mu, s^2)$，其中 $\mu = \mu(\theta)$ 且 $s^2 = s^2(\theta)$。那么 Z^2/s^2 服从自由度为 1、非中心参数为 $\lambda = (\mu/s)^2$ 的非中心卡方分布。因此，对于任意 $a > -1/2$,

$$\mathbb{E}\Big[e^{-a\,Z^2}\Big] = \frac{1}{\sqrt{1+2as^2}}\, \exp\!\left\{-\frac{\mu^2\, a}{1+2as^2}\right\}.$$

令 $a = (2\tau^2)^{-1}$，可得

$$\mathbb{E}\left[\exp\left\{-\frac{Z^2}{2\tau^2}\right\}\right] = \frac{1}{\sqrt{1+s^2/\tau^2}}\, \exp\left\{-\frac{\mu^2}{2(\tau^2+s^2)}\right\} = \sqrt{\frac{\tau^2}{\tau^2+s^2}}\, \exp\left\{-\frac{\mu^2}{2(\tau^2+s^2)}\right\},$$

由此得到所述表达式。 ■

证明 [引理 3 的证明。4]根据引理 3.3，对于每个 θ,

$$L_M(\theta;\tau) = \phi\big(s^2(\theta), \mu(\theta)\big), \qquad L_\infty(\theta;\tau) = \phi\big(0, \mu(\theta)\big),$$

其中，令 $t \geq 0$ 和 $m \in \mathbb{R}$，有

$$\phi(t,m) \;:=\; \sqrt{\frac{\tau^2}{\tau^2+t}}\, \exp\left\{-\frac{m^2}{2(\tau^2+t)}\right\}.$$

注意 $s^2(\theta) = v(\theta)/M \leq V/M$ 对所有 θ。

固定 θ，并在变量 t 上应用中值定理。对某个 $\xi \in (0, s^2(\theta))$,

$$\big|L_M(\theta;\tau) - L_\infty(\theta;\tau)\big| = \big|\phi\big(s^2(\theta), \mu(\theta)\big) - \phi\big(0, \mu(\theta)\big)\big| = \big|\partial_t \phi(\xi, \mu(\theta))\big|\, s^2(\theta).$$

我们现在关于 $t \in [0, V/M]$ 一致地给出 $\sup_{m\in\mathbb{R}} |\partial_t \phi(t,m)|$ 的上界。直接计算给出，其中 $c = \tau^2 + t$,

$$\partial_t \phi(t,m) = \frac{\tau}{2\, c^{3/2}}\, \exp\left\{-\frac{m^2}{2c}\right\}\, \Big(-1 + \frac{m^2}{c}\Big).$$

设 $u = m^2/c \geq 0$。则

$$\big|\partial_t \phi(t,m)\big| \leq \frac{\tau}{2\, c^{3/2}}\, \exp\Big(-\frac{u}{2}\Big)\, \big(1+u\big).$$

函数 $g(u) = (1+u)e^{-u/2}$ 在 $u \geq 0$ 处的最大值在 $u = 1$ with $g(1) = 2e^{-1/2}$处取得。因此

$$\sup_{m\in\mathbb{R}} \big|\partial_t \phi(t,m)\big| \leq \frac{\tau}{2\, c^{3/2}} \cdot 2e^{-1/2} = \frac{\tau e^{-1/2}}{(\tau^2+t)^{3/2}} \leq \frac{\tau e^{-1/2}}{\tau^3} = \frac{e^{-1/2}}{\tau^2},$$

因为 $t \geq 0$。因此，

$$\left|L_M(\theta;\tau) - L_\infty(\theta;\tau)\right| \leq \frac{e^{-1/2}}{\tau^2}\, s^2(\theta) \leq \frac{e^{-1/2}}{\tau^2}\, \frac{V}{M}.$$

对 $\theta \in \Theta$ 取上确界，得到

$$\sup_{\theta\in\Theta}\left|L_M(\theta;\tau) - L_\infty(\theta;\tau)\right| \leq \frac{e^{-1/2}V}{\tau^2 M} \xrightarrow[M\to\infty]{} 0,$$

这就证明了该结论以及一致的 $\mathcal{O}(1/M)$ 收敛速率。 ■

证明 [定理 3 的证明。6] 设 $U_j := \widetilde{w}_j\, h(\theta_j)$ 且 $V_j := \widetilde{w}_j$。给定 θ_j，批次平均值 R_{θ_j} 在各个 j 之间独立生成，因此 $\{(U_j,\ V_j)\}_{j\geq 1}$ 是独立同分布的（相对于先验分布 p 与模拟器随机性的乘积分布）。由强大数定律及可积性假设，

$$\frac{1}{n_\theta}\sum_{j=1}^{n_\theta} U_j \xrightarrow{\text{a.s.}} \mathbb{E}[U_1], \qquad \frac{1}{n_\theta}\sum_{j=1}^{n_\theta} V_j \xrightarrow{\text{a.s.}} \mathbb{E}[V_1],$$

其中 $\mathbb{E}[V_1] = \mathbb{E}[L_M(\Theta;\ \tau)] > 0$。因此，由连续映射定理，

$$\sum_{j=1}^{n_\theta} w_j\, h(\theta_j) = \frac{\sum_{j=1}^{n_\theta} U_j}{\sum_{j=1}^{n_\theta} V_j} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}[U_1]}{\mathbb{E}[V_1]}.$$

接下来只需确定极限。利用迭代期望，

$$\mathbb{E}[U_1] = \mathbb{E}[\mathbb{E}[\widetilde{w}_1\, h(\theta_1) \mid \theta_1]] = \mathbb{E}\Big[h(\Theta)\,\mathbb{E}\Big[e^{-R^2_{M,\Theta}/(2\tau^2)} \mid \Theta\Big]\Big] = \mathbb{E}[h(\Theta)\, L_M(\Theta;\tau)]\,,$$

类似地，$\mathbb{E}[V_1] = \mathbb{E}[L_M(\Theta;\ \tau)]$。因此，

$$\frac{\mathbb{E}[U_1]}{\mathbb{E}[V_1]} = \frac{\int h(\theta)\, L_M(\theta;\tau)\, p(\theta)\,\mathrm{d}\theta}{\int L_M(\theta;\tau)\, p(\theta)\,\mathrm{d}\theta} = \int h(\theta)\,\pi_{M,\tau}(\mathrm{d}\theta),$$

这就证明了该论断。 ■

证明 [定理 3.8 的证明。] 记 $Z_M := \int L_M(\theta;\ \tau)\ p(\theta)\,\mathrm{d}\theta$ 和 $Z_\infty := \int L_\infty(\theta;\ \tau)\ p(\theta)\,\mathrm{d}\theta$。由于 $0 \leq L_M,\ L_\infty \leq 1$，两者 $Z_M,\ Z_\infty \in (0,1]$ 且均有限。由引理 3.4，$\|L_M - L_\infty\|_\infty \to 0$，因此 $|Z_M - Z_\infty| \leq \|L_M - L_\infty\|_\infty \to 0$，进而 $Z_M \to Z_\infty > 0$。

假设 $|h| \leq K < \infty$。那么

$$\begin{aligned}\Phi_M(h) - \Phi_\infty(h) &= \frac{\int hL_M\,\mathrm{d}p}{Z_M} - \frac{\int hL_\infty\,\mathrm{d}p}{Z_\infty} \\ &= \frac{\int h(L_M - L_\infty)\,\mathrm{d}p}{Z_M} + \Big(\frac{1}{Z_M} - \frac{1}{Z_\infty}\Big)\int hL_\infty\,\mathrm{d}p,\end{aligned}$$

因此

$$|\Phi_M(h)-\Phi_\infty(h)| \le \frac{\int |h|\,|L_M-L_\infty|\,\mathrm{d}p}{Z_M} + \frac{|Z_M-Z_\infty|}{Z_M Z_\infty}\int |h|\,L_\infty\,\mathrm{d}p.$$

利用 $\int |h|\,|L_M-L_\infty|\,\mathrm{d}p \le K\,\|L_M-L_\infty\|_\infty$ 和 $\int |h|\,L_\infty\,\mathrm{d}p \le K Z_\infty$，得到

$$|\Phi_M(h)-\Phi_\infty(h)| \le \frac{K}{Z_M}\,\|L_M-L_\infty\|_\infty + \frac{K}{Z_M}\,|Z_M-Z_\infty| \le \frac{2K}{Z_M}\,\|L_M-L_\infty\|_\infty.$$

因此，当 $M\to\infty$ 时，$|\Phi_M(h)-\Phi_\infty(h)|\to 0$；此外，若 $\|L_M-L_\infty\|_\infty = \mathcal{O}(M^{-1})$，则同一式子给出有界 h 的 $\mathcal{O}(M^{-1})$ 速率。 ■

证明 [定理 3.13 的证明。] 定义随机映射

$$T(\beta) \;:=\; \frac{N(\beta)}{Z(\beta)}, \qquad N(\beta) := \int h(\theta)\,L_M^{(\beta)}(\theta;\tau)\,p(\theta)\,\mathrm{d}\theta, \quad Z(\beta) := \int L_M^{(\beta)}(\theta;\tau)\,p(\theta)\,\mathrm{d}\theta.$$

该结论是 $T(\widehat{\beta}) \xrightarrow{P} T(\beta^\circ)$。

对于任意实现 $\{(X_m,\ Y_m)\}_{m=1}^M$，映射 $\beta \mapsto \exp\{-\frac{1}{2\tau^2}\left(M^{-1}\sum_{m=1}^M (Y_m-\beta^\top \mathbf{X}_m)\right)^2\}$是连续的，且有界于 1。由于 $(X_m,\ Y_m)$ 的模拟器分布不依赖于β，根据控制收敛定理，$L_M^{(\beta)}(\theta;\ \tau)$ 对每个固定的 θ 都关于 β 连续。

根据假设 3.12(ii)，存在一个可积包络 g，使得对 β° 的某个邻域 $\mathcal{N}$ 中的所有 β，均有 $L_M^{(\beta)}(\theta;\ \tau) \le g(\theta)$。由于 h 和 $L_M^{(\beta)}(\theta;\ \tau)$ 有界，对 θ 应用控制收敛可得，当 $\beta\to\beta^\circ$ 时，$N(\beta)\to N(\beta^\circ)$ 且 $Z(\beta)\to Z(\beta^\circ)$。此外，$0 < L_M^{(\beta)} \le 1$ 意味着 $Z(\beta^\circ)\in(0,1]$。

由于 $Z(\beta^\circ) > 0$ 且 $Z(\beta)\to Z(\beta^\circ)$，存在一个邻域，使得其中的 $Z(\beta)$ 始终远离零；因此，$T(\beta)$ 在 β° 处连续。由于 $\widehat{\beta}\xrightarrow{P}\beta^\circ$ 且 T 在 β° 处连续，依据连续映射定理可得 $T(\widehat{\beta})\xrightarrow{P} T(\beta^\circ)$，即

$$\int h(\theta)\,\pi_{M,\tau}^{(\widehat{\beta})}(\mathrm{d}\theta) \;\xrightarrow{P}\; \int h(\theta)\,\pi_{M,\tau}^{(\beta^\circ)}(\mathrm{d}\theta).$$

■

证明 [引理 4.4 的证明。] 由于 β° 是在 $\mathcal{L}_{\mathrm{obs}}$下通过带截距的 OLS 投影得到的，正规方程包含截距方程 $\mathbb{E}_{\mathrm{obs}}\big[Y-\beta^{\circ\top}\mathbf{X}\big]=0$。如果$\mathcal{L}_{\mathrm{sim}}(\cdot\mid\theta^\star)=\mathcal{L}_{\mathrm{obs}}$，那么

$$\mu^\circ(\theta^\star) = \mathbb{E}_{\mathrm{sim}(\theta^\star)}\big[Y-\beta^{\circ\top}\mathbf{X}\big] = \mathbb{E}_{\mathrm{obs}}\big[Y-\beta^{\circ\top}\mathbf{X}\big] = 0.$$

因此 $\mu^\circ(\theta^\star)^2=0$，由此得到 $\inf_{\vartheta\in\Theta}\mu^\circ(\vartheta)^2=0$ 和 $\theta^\star\in\Theta^\dagger$。 ■

证明 [引理 4 的证明。5] 固定 $\theta\in\Theta$令 Z_m, $m=1,\dots,M$ 为 Z 的独立同分布副本（条件于 θ），并定义

$$R_{M,\theta} = \frac{1}{M}\sum_{m=1}^M Z_m\ , \qquad S_{M,\theta} = \frac{\sqrt{M}\,(R_{M,\theta}-\mu^\circ(\theta))}{\sqrt{v^\circ(\theta)}}.$$

因此，$S_{M,\theta}$ 的均值为 0，方差为 1（条件于 θ）。令 $\sigma(\theta) = \sqrt{v^\circ(\theta)}$ 和 $\widetilde{Z} := Z - \mu^\circ(\theta)$，使得 $\mathbb{E}[\widetilde{Z} \mid \theta] = 0$，**方差**$(\widetilde{Z} \mid \theta) = v^\circ(\theta)$，且 $S_{M,\theta} = M^{-1/2} \sum_{m=1}^{M} \widetilde{Z}_m / \sigma(\theta)$。令

$$\rho_3(\theta) := \mathbb{E}\left[\left|\frac{\widetilde{Z}}{\sigma(\theta)}\right|^3 \middle| \theta\right].$$

我们证明 $\sup_{\theta\in\Theta} \rho_3(\theta) < \infty$。由三角不等式和初等界 $|a-b|^3 \le 4(|a|^3 + |b|^3)$，

$$\mathbb{E}[|\widetilde{Z}|^3 \mid \theta] \;\le\; 4\Big(\mathbb{E}[|Z|^3 \mid \theta] + |\mu^\circ(\theta)|^3\Big).$$

回顾 $\mathbb{E}[|Z|^3 \mid \theta] \le C$ 在 θ上一致成立。μ° 在紧集 Θ 上的连续性意味着 $\sup_\theta |\mu^\circ(\theta)| < \infty$。因此，

$$\sup_{\theta\in\Theta} \mathbb{E}[|\widetilde{Z}|^3 \mid \theta] \;<\; \infty, \qquad \text{and hence} \qquad \sup_{\theta\in\Theta} \rho_3(\theta) \;\le\; \frac{\sup_\theta \mathrm{E}[|\widetilde{Z}|^3 \mid \theta]}{v_{\min}^{3/2}} \;<\; \infty.$$

随后，对独立同分布和适用（均匀）Berry–Esseen不等式，从而存在一个普适常数 C_{BE}，使得对所有 $t \in \mathbb{R}$ 和所有 $\theta \in \Theta$，

$$|\Pr(S_{M,\theta} \le t \mid \theta) - \Phi(t)| \;\le\; \frac{C_{\mathrm{BE}}\, \rho_3(\theta)}{\sqrt{M}} \;\le\; \frac{C'}{\sqrt{M}},$$

其中 $C' := C_{\mathrm{BE}} \sup_\theta \rho_3(\theta) < \infty$。最后，由于

$$\frac{R_{M,\theta} - \mu^\circ(\theta)}{\sqrt{v^\circ(\theta)/M}} \;=\; S_{M,\theta},$$

该式恰好就是所断言的界。对 $t \in \mathbb{R}$ 和 $\theta \in \Theta$取上确界，即完成证明。

■

证明 [引理 4.6 的证明。] 固定 $\theta \in \Theta$，并简记 $\mu = \mu^\circ(\theta)$ 和 $s^2 = v^\circ(\theta)/M$。回顾

$$L_M(\theta;\tau) = \mathbb{E}\left[\exp\left\{-\frac{R_{M,\theta}^2}{2\tau^2}\right\}\middle|\theta\right], \qquad R_{M,\theta} = \frac{1}{M}\sum_{m=1}^{M}\big(Y_m^{\mathrm{sim}} - \beta^{\circ\top}\mathbf{X}_m^{\mathrm{sim}}\big).$$

令 F_θ 表示给定 θ 时 R_M 的累积分布函数，θ，并令 G_θ 表示 $\mathcal{N}(\mu, s^2)$ 的累积分布函数。根据引理 4.5，

$$\Delta_M \;:=\; \sup_{t\in\mathbb{R}} |F_\theta(t) - G_\theta(t)| \;\le\; \frac{C'}{\sqrt{M}}, \tag{39}$$

其中 C' 与 θ相互独立。

为控制真实分布与高斯分布下期望之差，考虑测试函数 $g(r) = \exp(-r^2/(2\tau^2))$。由于 g 绝对连续且导数有界，我们可以将期望之差表示为关于分布函数 F_θ和 G_θ的斯蒂尔切斯积分，其中它们分别对应于 $R_{M,\,\theta}$和 $\mathcal{N}(\mu,\ s^2)$:

$$\mathbb{E}[g(R_{M,\theta})] - \mathbb{E}_{Z\sim\mathcal{N}(\mu,s^2)}[g(Z)] \;=\; \int_{\mathbb{R}} g(r)\,\mathrm{d}F_\theta(r) - \int_{\mathbb{R}} g(r)\,\mathrm{d}G_\theta(r) \;=\; \int_{\mathbb{R}} g(r)\,\mathrm{d}\big(F_\theta(r) - G_\theta(r)\big).$$

根据有界变差函数的分部积分公式（例如，参见 Shorack 和 Wellner 的 1986年版定理 13.2.1），对于任意两个右连续、单调不减且有界变差的函数 F，G，以及任意在 $\mathbb{R}$上导数 g' 可积的绝对连续函数 g，

$$\left| \int g \, \mathrm{d}(F - G) \right| \;\le\; \mathrm{Var}(g) \, \sup_{t \in \mathbb{R}} |F(t) - G(t)|,$$

其中，$\mathrm{Var}(g) = \int_{\mathbb{R}} |g'(r)| \mathrm{d}r$ 表示 g的全变差。在我们的情形中，$F = F_\theta$ 和 $G = G_\theta$是分布函数，因此，上确界 $\sup_{t \in \mathbb{R}} |F(t) - G(t)|$ 等于柯尔莫哥洛夫距离 Δ_M。由于 $g'(r) = -(r/\tau^2)\exp(-r^2/(2\tau^2))$，其全变差有限且易于计算：

$$\mathrm{Var}(g) = \int_{-\infty}^{\infty} |g'(r)| \, \mathrm{d}r = \frac{1}{\tau^2} \int_{-\infty}^{\infty} |r| \, e^{-r^2/(2\tau^2)} \, \mathrm{d}r = \frac{2}{\tau^2} \int_0^{\infty} r \, e^{-r^2/(2\tau^2)} \, \mathrm{d}r = 2.$$

结合引理 4.5 中的 Berry–Esseen界，

$$\Delta_M = \sup_t |F_\theta(t) - G_\theta(t)| \le \frac{C'}{\sqrt{M}},$$

可得一致不等式

$$\Big| \mathbb{E}[g(R_{M,\theta})] - \mathbb{E}_{Z \sim \mathcal{N}(\mu, s^2)}[g(Z)] \Big| \;\le\; \mathrm{Var}(g) \, \Delta_M \;\le\; 2 \, \frac{C'}{\sqrt{M}}, \tag{40}$$

该不等式对所有 $\theta \in \Theta$均成立。由于在假设 4.1 下，$\mathrm{Var}(g) = 2$ 和 Berry–Esseen 常数 C' 均独立于 θ，因此此界关于 θ 是一致的。

剩下的工作是精确计算高斯期望。根据引理 3.3，对于 $Z \sim \mathcal{N}(\mu, s^2)$，

$$\mathbb{E}\left[\exp\left\{ -\frac{Z^2}{2\tau^2} \right\} \right] = \sqrt{\frac{\tau^2}{\tau^2 + s^2}} \, \exp\left\{ -\frac{\mu^2}{2(\tau^2 + s^2)} \right\}.$$

代入 $s^2 = v^\circ(\theta)/M$，即可得到公式（25）中的首项，其中 $\tau^2_{\mathrm{eff}}(\theta;\, M,\, \tau) = \tau^2 + v^\circ(\theta)/M$。最后，定义

$$r_{M,\tau}(\theta) \;:=\; \mathbb{E}[g(R_{M,\theta})] - \mathbb{E}_{Z \sim \mathcal{N}(\mu, s^2)}[g(Z)].$$

界公式（40）表明 $\sup_{\theta \in \Theta} |r_{M,\,\tau}(\theta)| \le 2C'/\sqrt{M}$。令 $C'' = 2C'$即完成证明。常数 C'，C'' 关于 θ 一致，且不依赖于 τ。 ∎

证明 [引理 4.7 的证明。] 令 $m := \inf_{\vartheta \in \Theta} \mu^\circ(\vartheta)^2$。根据假设 4.2，引理 4.4 表明 $m = 0$。根据假设 4.3，对于每个 $\theta \notin U$，我们都有 $\mu^\circ(\theta)^2 \ge m + \eta = \eta$。令 $s^2(\theta) := v^\circ(\theta)/M \in [v_{\min}/M,\, v_{\max}/M]$。根据引理 4.6，

$$L_M(\theta;\tau) = \sqrt{\frac{\tau^2}{\tau^2 + s^2(\theta)}} \, \exp\left\{ -\frac{\mu^\circ(\theta)^2}{2(\tau^2 + s^2(\theta))} \right\} + r_{M,\tau}(\theta), \qquad \sup_{\theta \in \Theta} |r_{M,\tau}(\theta)| \le \frac{C''}{\sqrt{M}}. \tag{41}$$

分子界（在 U 之外）。利用 $\sqrt{\tau^2+s^2(\theta)}\le 1$ 和 $s^2(\theta) \le v_{\max}/M$，

$$L_M(\theta;\tau) \;\le\; \exp\!\left\{-\frac{\mu^\circ(\theta)^2}{2(\tau^2+v_{\max}/M)}\right\} + \frac{C''}{\sqrt{M}} \;\le\; \exp\!\left\{-\frac{\eta}{2(\tau^2+v_{\max}/M)}\right\} + \frac{C''}{\sqrt{M}}.$$

积分后，

$$\int_{\Theta\setminus U} p(\theta)\, L_M(\theta;\tau)\,\mathrm{d}\theta \;\le\; A\,\exp\!\left\{-\frac{\eta}{2(\tau^2+v_{\max}/M)}\right\} + \frac{C''}{\sqrt{M}}, \tag{42}$$

其中 $A := p_{\max}(\Theta)\lambda(\Theta)$，$p_{\max}(\Theta) := \sup_{\theta\in\Theta} p(\theta)$，且 $\lambda(\Theta) \in (0,\infty)$ 是紧集 Θ 的勒贝格测度。

分母界（在 U 内）。由于 $U \supset \Theta^\dagger$ 是开集且 μ° 连续，因此存在一个内部非空的紧集 $V \subset U$，并且对任意 $\epsilon > 0$，

$$\mu^\circ(\theta)^2 \;\le\; m + \epsilon = \epsilon \qquad \forall\,\theta \in V. \tag{43}$$

Fix

$$\epsilon \;:=\; \frac{\tau^2+v_{\min}/M}{\tau^2+v_{\max}/M}\,\frac{\eta}{2} \;\le\; \frac{\eta}{2}.$$

利用方程（41），$s^2(\theta) \in [v_{\min}/M,\, v_{\max}/M]$，以及方程（43），

$$L_M(\theta;\tau) \;\ge\; \sqrt{\frac{\tau^2}{\tau^2+v_{\max}/M}}\,\exp\!\left\{-\frac{\epsilon}{2(\tau^2+v_{\min}/M)}\right\} - \frac{C''}{\sqrt{M}} \qquad \forall\,\theta \in V.$$

由于 p 在 $\Theta^\dagger$ 的一个邻域上连续且严格为正（假设 4.3），存在 $p_{\min}(B) > 0$，使得在 V 上 $p(\theta) \ge p_{\min}(V)$。令 $\lambda(V)$ 为紧集 $V \subset U$ 的勒贝格测度。在 V 上积分可得，其中 $B := p_{\min}(V)\lambda(V)$，

$$\int_U p(\theta)\, L_M(\theta;\tau)\,\mathrm{d}\theta \;\ge\; B\sqrt{\frac{\tau^2}{\tau^2+v_{\max}/M}}\,\exp\!\left\{-\frac{\epsilon}{2(\tau^2+v_{\min}/M)}\right\} - \frac{C''}{\sqrt{M}}. \tag{44}$$

根据 ϵ 的选取，

$$-\frac{\eta}{2(\tau^2+v_{\max}/M)} + \frac{\epsilon}{2(\tau^2+v_{\min}/M)} \;=\; -\frac{\eta}{4(\tau^2+v_{\max}/M)}.$$

对于足够大的 M以及 $M\tau^2 \ge K$（针对下文选取的某个 K），方程（44）的右端为正，且因子 $\sqrt{\frac{\tau^2}{\tau^2+v_{\max}/M}}$ 以某个常数 c_K 为下界（$0<c_K<1$；事实上，$M\tau^2 \ge K$ 蕴含 $\tau^2/(\tau^2+v_{\max}/M) \ge K/(K+v_{\max}) =: c_K^2$。综合方程（42）–（44）并对指数项作因式分解，可得

$$\frac{\int_{\Theta\setminus U} pL_M}{\int_U pL_M} \;\le\; \frac{A}{B\,c_K}\,\exp\!\left\{-\frac{\eta}{4(\tau^2+v_{\max}/M)}\right\}\,\frac{1+\frac{C''}{\sqrt{M}A}\,e^{\frac{\eta}{2(\tau^2+v_{\max}/M)}}}{1-\frac{C''}{\sqrt{M}Bc_K}\,e^{\frac{\epsilon}{2(\tau^2+v_{\min}/M)}}}.$$

对于 $M \ge M_0$ （足够大，使得最后一个分式的分母超过 $1/2$），对某个仅依赖于 $C'', A, B, c_K, \eta, v_{\min}, v_{\max}$ 的 C'''，最后一个分式以 $1+C'''/\sqrt{M}$ 为界，并且我们可以令 $C(U) := (A/(B\,c_K)) \cdot 2$，从而得出

$$\frac{\int_{\Theta\setminus U} pL_M}{\int_U pL_M} \;\le\; C(U)\,\exp\!\Big\{-\frac{\eta}{4(\tau^2 + v_{\max}/M)}\Big\}\Big(1+\frac{C'''}{\sqrt{M}}\Big).$$

一旦固定 K 和 M_0，所有常数都与 $(M,\ \tau)$ 无关。 ■

证明 [定理 4.8 的证明。] 由矩可表示性和引理 4.4，$m := \inf_{\vartheta\in\Theta}\mu^\circ(\vartheta)^2 = 0$。固定一个开集 $U \supset \Theta^\dagger$，并令 $\eta = \eta(U) > 0$ 为假设 4.3 中的分离常数。由总体拉普拉斯界引理 4.7，存在常数 $C(U)$ $C''' < \infty$ 和 $M_0,\ K > 0$（与 $M,\ \tau$ 无关），使得对所有 $M \ge M_0$ 以及所有满足 $M\tau^2 \ge K$ 的 $\tau > 0$，

$$\frac{\int_{\Theta\setminus U} p(\theta)L_M(\theta;\tau)\,\mathrm{d}\theta}{\int_U p(\theta)L_M(\theta;\tau)\,\mathrm{d}\theta} \;\le\; C(U)\,\exp\!\Big\{-\frac{\eta}{4(\tau^2 + v_{\max}/M)}\Big\}\Big(1+\frac{C'''}{\sqrt{M}}\Big). \tag{45}$$

条件 (26) 意味着对所有充分大的 M，$M\tau_M^2 \ge K$，因此方程 (45) 可应用于 $\tau = \tau_M$，且右侧在 $M\to\infty$ 和 $\tau_M \downarrow 0$ 时趋于 0。由于

$$\pi_{M,\tau_M}(U) = \frac{\int_U pL_M}{\int_\Theta pL_M} = \frac{1}{1 + \frac{\int_{\Theta\setminus U} pL_M}{\int_U pL_M}},$$

我们得到 $\pi_{M,\ \tau_M}(U) \to 1$。对于一个具有 $F \cap \Theta^\dagger = \emptyset$ 的闭 F，选取一个具有 $\Theta^\dagger \subset U \subset \overline{U} \subset F^c$ 的开 U；于是 $\pi_{M,\ \tau_M}(F) \le 1 - \pi_{M,\ \tau_M}(U) \to 0$。这证明了总体结论。

对于经验测度，令 A 为任意博雷尔集（特别是 $A = U$ 或 $A = F$）。由定理 3.6，对于每个固定的 M，

$$\Pi_{n_\theta}(A) = \sum_{j=1}^{n_\theta} w_j\,\mathbb{1}_A(\theta_j) \;\xrightarrow[n_\theta\to\infty]{\text{a.s.}}\; \pi_{M,\tau_M}(A).$$

给定 $\varepsilon > 0$，首先取足够大的 M，使得 $\pi_{M,\ \tau_M}(U) \ge 1-\varepsilon/2$ 且 $\pi_{M,\ \tau_M}(F) \le \varepsilon/2$；然后取足够大的 $N(M)$，使得对所有 $n_\theta \ge N(M)$，$|\Pi_{n_\theta}(A) - \pi_{M,\ \tau_M}(A)| \le \varepsilon/2$ 几乎必然成立。因此，$\Pi_{n_\theta}(U) \ge 1-\varepsilon$ 且 $\Pi_{n_\theta}(F) \le \varepsilon$ 几乎必然成立。令 $\varepsilon \downarrow 0$，即可得到所述的几乎必然收敛。

■

证明 [命题 4 的证明。9] 令 β° 表示在观测分布 $\mathcal{L}_{\text{obs}}(X,Y)$ 下，Y 在增广预测变量 $\mathbf{X} = (1, X^\top)^\top$ 上的最佳线性投影。带截距的普通最小二乘拟合得到

$$\mathbb{E}_{\text{obs}}\big[\mathbf{X}\,(Y - \beta^{\circ\top}\mathbf{X})\big] = \mathbf{0}.$$

取第一个（截距）分量可得 $\mathbb{E}_{\text{obs}}[\,Y - \beta^{\circ\top}\mathbf{X}\,] = 0$。

固定 $\theta^\star \in \Theta^\star$。分布的精确相等性 $\mathcal{L}_{\mathrm{sim}}(X, Y \mid \theta^\star) = \mathcal{L}_{\mathrm{obs}}(X, Y)$，推出对于所有可积的 φ，有 $\mathbb{E}_{\mathrm{sim}}[\varphi(X, Y) \mid \theta^\star] = \mathbb{E}_{\mathrm{obs}}[\varphi(X, Y)]$。特别地，

$$\mu^\circ(\theta^\star) = \mathbb{E}_{\mathrm{sim}}\big[Y - \beta^{\circ\top}\mathbf{X} \mid \theta^\star\big] = \mathbb{E}_{\mathrm{obs}}\big[Y - \beta^{\circ\top}\mathbf{X}\big] = 0.$$

因此，$\mu^\circ(\theta^\star)^2$ 达到全局最小值 0，所以 $\theta^\star \in \arg\min_{\theta\in\Theta} \mu^\circ(\theta)^2 = \Theta^\dagger$。因此 $\Theta^\star \subseteq \Theta^\dagger$。

■

参考文献

M. A. Beaumont。进化与生态学中的近似贝叶斯计算。《生态学、进化与系统学年度评论》，41(1):379–406，2010年。P. Behroozi、R. H. Wechsler、A. P. Hearin 和 C. Conroy。UNIVERSEMACHINE：从 z $= 0$-10 探究星系增长与暗物质晕组装之间的相关性。《皇家天文学会月报》，488(3):3143–3194，2019年9月。doi: 10.1093/ mnras/stz1182。A. Bharti、D. Huang、S. Kaski 和 F.-X. Briol。成本感知的基于模拟的推断。载于《人工智能与统计国际会议》，第28–36页。PMLR，2025年。P. Billingsley。《概率与测度》。约翰·威立父子出版公司，2017年。P. G. Bissiri、C. C. Holmes 和 S. G. Walker。更新信念分布的通用框架。《皇家统计学会杂志B辑：统计方法学》，78(5):1103–1130，2016。M. G. Blum、M. A. Nunes、D. Prangle 和 S. A. Sisson。近似贝叶斯计算中的降维方法比较综述。《统计科学》，第189–208页，2013年。J. Brehmer。粒子物理学中的基于模拟的推断。《自然综述：物理学》，3(5):305–305，2021年。V. Busillo、C. Tortora、N. R. Napolitano、L. V. Koopmans、G. Covone、F. Gentile 和 L. K. Hunt。Casco：基于宇宙学模拟与观测数据的宇宙学和天体物理参数——I：约束本地恒星形成星系中的物理过程。《皇家天文学会月报》，525(4):6191–6213，2023年。V. Busillo、C. Tortora、G. Covone、L. V. Koopmans、M. Silvestrini 和 N. R. Napolitano。Casco：基于宇宙学模拟与观测数据的宇宙学和天体物理参数——II：利用早型和晚型星系约束宇宙学与天体物理过程。《天文学与天体物理学》，693:A112，2025年。K. Cranmer、J. Pavez 和 G. Louppe。利用校准的判别分类器近似似然比。arXiv预印本 arXiv:1506.02169，2015年。

K. Cranmer、J. Brehmer 和 G. Louppe。基于模拟的推断前沿。美国国家科学院院刊，117(48):30055–30062，2020年。K. Csill´ery、M. G. Blum、O. E. Gaggiotti 和 O. Fran¸cois。实践中的近似贝叶斯计算（近似贝叶斯计算）。生态学与进化趋势，25(7):410–418，2010年。C. C. Drovandi。近似贝叶斯计算与间接推断。载于《近似贝叶斯计算手册》，第179–209页。Chapman and Hall/CRC，2018年。C. C. Drovandi、A. N. Pettitt 和 A. Lee。使用参数辅助模型进行贝叶斯间接推断。《统计科学》，30(1):72–95，2015年。A. Farahi、V. Guglielmo、A. E. Evrard、B. M. Poggianti、C. Adami、S. Ettori、F. Gastaldello、P. A. Giles、B. J. Maughan、D. Rapetti 等。XXL巡天-XXIII：通过集合光谱学确定XXL星系团的质量尺度。《天文学与天体物理学》，620:A8，2018年。P. Fearnhead 和 D. Prangle。为近似贝叶斯计算构建汇总统计量：半自动近似贝叶斯计算。《皇家统计学会杂志B辑：统计方法学》，74(3):419–474，2012年。F. Forbes、H. D. Nguyen、T. Nguyen 和 J. Arbel。利用替代后验分布进行近似贝叶斯计算的汇总统计量与差异度量。《统计学与计算》，32(5):85，2022。D. T. Frazier、G. M. Martin、C. P. Robert 和 J. Rousseau。近似贝叶斯计算的渐近性质。《生物统计学》，105(3):593–607，2018年。A. M. Garcia、J. C. Rose、P. Torrey、A. Caputo、M. Lisanti、A. B. Pace、H. Liu、A. Hussein、H. Liu、F. Villaescusa-Navarro 等。DREAMS项目：剖析晕间方差与重子反馈对银河系暗物质密度轮廓的影响。arXiv预印本 arXiv:2512.03132，2025年。D. Greenberg、M. Nonnenmacher 和 J. Macke。用于无似然推断的自动后验变换。载于《国际机器学习会议》，第2404–2414页。PMLR，2019年。J. E. Greene、J. Strader 和 L. C. Ho。中等质量黑洞。《天文学与天体物理学年度评论》，58:257–312，2020年8月。doi: 10.1146/ annurev-astro-032620-021835。H. Heesterbeek、R. M. Anderson、V. Andreasen、S. Bansal、D. De Angelis、C. Dye、K. T. Eames、W. J. Edmunds、S. D. Frost、S. Funk 等。全球卫生复杂图景中的传染病动力学建模。《科学》，347(6227):aaa4339，2015年。K. Heggland 和 A. Frigessi。间接推断中的估计函数。《皇家统计学会杂志B辑：统计方法学》，66(2):447–462，2004年。J. Hermans、V. Begy 和 G. Louppe。使用摊销近似比率估计器进行无似然MCMC。载于《国际机器学习会议》，第4239–4248页。PMLR，2020年。

J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy 和 G. Louppe。基于模拟的推断中的信任危机？你的后验近似可能并不忠实。arXiv预印本 arXiv:2110.06581，2021年。D. Huang, A. Bharti, A. Souza, L. Acerbi 和 S. Kaski。模型设定错误时，学习用于基于模拟的推断的稳健统计量。《神经信息处理系统进展》，36：7289–7310，2023年。D. Huterer 和 D. L. Shafer。暗能量二十年后：可观测量、探测手段、一致性检验。《物理学进展报告》，81(1)：016901，2017年。R. Jiang, P. Y. Lu 和 R. Willett。嵌入与模拟：用于基于模拟的推断的对比表征。arXiv预印本 arXiv:2409.18402，2024年。H. Kitano。计算系统生物学。《自然》，420(6912)：206–210，2002年。C.-Y. Lai, P. Hassanzadeh, A. Sheshadri, M. Sonnewald, R. Ferrari 和 V. Balaji。气候物理学与模拟中的机器学习。《凝聚态物理年度评论》，16(1)：343–365，2025年。E. Lilie, J. C. Rose, M. Lisanti, A. M. Garcia, P. Torrey, K. E. Kollmann, J. Li, O. Mostow, B. Y. Wang, S. O'Neil 等。DREAMS项目：厘清晕间方差和重子反馈对银河系暗物质速度分布的影响。arXiv预印本 arXiv:2512.04157，2025年。J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski 和 J. Corander。近似贝叶斯计算的基础与最新进展。《系统生物学》，66(1)：e66–e82，2017年。J.-M. Lueckmann, G. Bassetto, T. Karaletsos 和 J. H. Macke。借助模拟器网络进行无似然推断。载于《近似贝叶斯推断进展研讨会》，第32–53页。PMLR，2019年。J.-M. Marin, P. Pudlo, C. P. Robert 和 R. J. Ryder。近似贝叶斯计算方法。《统计学与计算》，22(6)：1167–1180，2012年。R. Martin 和 N. Syring。直接对风险最小化解进行吉布斯后验推断：构建、集中性与校准。载于《统计学手册》，第47卷，第1–41页。爱思唯尔，2022年。J. W. Miller。广义后验的渐近正态性、集中性与覆盖率。《机器学习研究杂志》，22(168)：1–53，2021年。T. Nguyen, F. Villaescusa-Navarro, S. Mishra-Sharma, C. Cuesta-Lazaro, P. Torrey, A. Farahi, A. M. Garcia, J. C. Rose, S. O'Neil, M. Vogelsberger 等。梦想如何实现：利用扩散模型和点云模拟卫星星系与子晕种群。arXiv预印本 arXiv:2409.02980，2024年。G. Papamakarios 和 I. Murray。利用贝叶斯条件密度估计，对模拟模型进行快速的 ε-free 推断。《神经信息处理系统进展》，29，2016。

A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, L. Hernquist, P. Torrey, M. Vogelsberger, R. Weinberger 和 F. Marinacci。使用IllustrisTNG模型模拟星系形成。《皇家天文学会月报》，473 (3)：4077–4106，2018年1月。doi：10.1093/mnras/stx2656。C. P. Robert。贝叶斯计算工具。《统计学及其应用年度评论》，1(1)：153–177，2014年。J. C. Rose, M. Lisanti, P. Torrey, F. Villaescusa-Navarro, A. M. Garcia, A. Farahi, C. Filion, A. M. Brooks, N. Kallivayalil, K. E. Kollmann 等。DREAMS项目：厘清晕间方差和重子反馈对银河系卫星星系的影响。arXiv预印本 arXiv:2512.02095，2025年a。J. C. Rose, M. Lisanti, P. Torrey, F. Villaescusa-Navarro, A. M. Garcia, A. Farahi, C. Filion, A. M. Brooks, N. Kallivayalil, K. E. Kollmann 等。DREAMS项目：一套用于构建银河系背景并评估物理不确定性的全新1024个模拟。《天体物理学杂志》，arXiv预印本 arXiv:2512.00148，2025b。J. C. Rose, P. Torrey, F. Villaescusa-Navarro, M. Lisanti, T. Nguyen, S. Roy, K. E. Kollmann, M. Vogelsberger, F.-Y. Cyr-Racine, M. V. Medvedev 等。介绍DREAMS项目：结合机器学习与模拟研究暗物质和天体物理学。《天体物理学杂志》，982(2)：68，2025年c。S. A. Sisson, Y. Fan 和 M. Beaumont。《近似贝叶斯计算手册》。CRC出版社，2018年。M. Steinmetz 和 J. F. Navarro。塔利-费舍尔关系的宇宙学起源。《天体物理学杂志》，513(2)：555，1999年。S. Tavar´e, D. J. Balding, R. C. Griffiths 和 P. Donnelly。根据DNA序列数据推断合并时间。《遗传学》，145(2)：505–518，1997年。L. Tortorelli, S. Fischbacher, D. Gr¨un, A. Refregier, S. Bellstedt, A. S. Robotham 和 T. Kacprzak。Galsbi-sps：一种基于恒星族群合成、用于宇宙学和星系演化应用的星系族群模型。arXiv预印本 arXiv:2505.21610，2025年。M. Vogelsberger, F. Marinacci, P. Torrey 和 E. Puchwein。星系形成的宇宙学模拟。《自然综述：物理学》，2(1)：42–66，2020年。X. Wang, R. P. Kelly, A. L. Jenner, D. J. Warne 和 C. Drovandi。计算生物学中基于模拟的推断综合指南。arXiv预印本 arXiv:2409.19675，2024年。D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess 和 E. Rozo。宇宙加速膨胀的观测探针。《物理学报告》，530(2)：87–255，2013年。R. Weinberger, V. Springel, R. Pakmor, D. Nelson, S. Genel, A. Pillepich, M. Vogelsberger, F. Marinacci, J. Naiman, P. Torrey 和 L. Hernquist。IllustrisTNG模拟中的超大质量黑洞及其反馈效应。《皇家天文学会月报》，479(3)：4056–4072，2018年9月。doi：10.1093/mnras/sty1733。

A. Zammit-Mangion、M. Sainsbury-Dale 和 R. Huser。《摊销推断的神经方法》。《统计学及其应用年度评论》，12(1):311–335，2025年。