arXiv:2602.02216v1 [stat.ME] 2 Feb 2026

贝叶斯自助法程序中目标参数的后验不确定性

Magid Sabbagh
麦吉尔大学数学与统计学系，加拿大魁北克省蒙特利尔
magid.sabbagh@mail.mcgill.ca

David A. Stephens*麦吉尔大学数学与统计学系，加拿大魁北克省蒙特利尔 david.stephens@mcgill.ca

2026年2月3日

摘要

我们提出一种通用方法，可在存在有限维干扰参数的情况下，对有限维“目标”参数进行有效的贝叶斯分析。我们将该方法应用于基于估计方程的因果推断。尽管贝叶斯因果推断领域的大量文献采用传统的“似然乘以先验分布”框架，但近期提出的一种方法——“关联贝叶斯自助法”——偏离了这一经典框架，转而利用狄利克雷过程和贝叶斯自助法进行有效的贝叶斯推断。这些方法基于倾向评分进行调整，并说明了在研究处理效应的后验分布时，如何处理与倾向评分相关的不确定性。我们从理论上考察了所得后验分布的渐近性质，并证明，我们提出的方法——“关联贝叶斯自助法”的一种推广形式——具有理想的频率学派性质。此外，我们证明可信区间在渐近意义下具有正确的覆盖性质。我们还讨论了该方法在因果推断中模型设定错误及单重稳健模型下的应用。

关键词 因果I　　断；狄利克雷过程；贝叶斯自助法；关联贝叶斯自助法；倾向评分

1 引言

在推断问题中，关注点通常放在某个量上，该量就可观测量而言具有某种有意义的解释。在经典贝叶斯表示中，这类量通常是由德·芬尼蒂表示所定义的参数。然而，我们可以采取一种更一般的方法来“针对”参数，即将它们定义为数据生成机制的泛函，例如作为损失最小化问题的解而得到。这种更一般的目标参数定义并不总是适用于常规的贝叶斯推断，因为此类参数可能不会以直接的方式出现在似然函数中。通常，此类目标参数并不能完全确定数据生成过程。我们采用“目标参数”这一术语，而非更为常见的“感兴趣参数”，因为我们认为它能更好地反映推断目标。

当我们意识到模型可能存在设定错误时，也会涉及目标参数。例如，在临床研究中，我们可能关注一个用于量化处理对结果影响的参数。在这一因果推断场景中，Rosenbaum and Rubin（1983）强调了倾向评分在观察性研究中发挥的关键作用。基于倾向评分的调整可以通过加权、分层、匹配或回归来实现。尽管频率学派文献在研究因果效应时普遍采用拟合得到的倾向评分，但对于能否使用这种插件法进行完全贝叶斯推断，尚未达成普遍共识。Stephens 等人（2023）探讨了在可能存在模型设定错误的情况下，如何在基于狄利克雷过程的贝叶斯因果分析中纳入参数化建模的倾向评分。尽管对于基于拟合倾向评分值的两步法能否视为完全贝叶斯方法，尚无共识，但他们提出的计算方法

*通讯作者

能够对因果靶参数进行有效的贝叶斯推断。本文旨在为这一计算方法提供理论依据，并证明其所得后验分布确实渐近地具有名义水平的覆盖率。我们采用一个综合框架，运用经验过程、贝叶斯自助法和估计方程的相关思想。

在第2节中，我们概述一种源自狄利克雷过程与贝叶斯自助法应用的自助法。在第3节中，我们说明如何通过估计方程来确定目标参数，并利用这些概念来表述如下问题：在存在已估计干扰参数的情况下，如何获得目标参数的完全有效的贝叶斯推断。在第4节中，我们推导本文的主要理论结果。我们研究处理有关干扰参数的推断不确定性对目标参数后验分布的影响，并从理论上证明一种近期提出的方法——关联贝叶斯自助法——的有效性。接下来，我们给出一个论证，以验证在因果结构模型中基于倾向评分的两步程序可作为贝叶斯方法使用。在第5节中，我们通过使用G-估计和逆加权估计量对处理效应进行分析，实证验证先前获得的结果。

2 自助法与损失最小化

下文中我们采用如下记号：设 $O_1, \ldots, O_n$ 为定义在概率空间（$\mathcal{O}, B$）上、分布为 $F_0 = P_O$的独立同分布随机变量。将经验过程P_n 与加权经验过程P_n^w 记为并定义为

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{O_i} \qquad \text{and} \qquad \mathbb{P}_n^w = \sum_{i=1}^{n} w_{in}\delta_{O_i}$$

分别地，其中 $\{w_{in}, i = 1, \ldots, n\}$ 是一组总和为一的（可能是随机的）权重。本文仅关注狄利克雷权重。

2.1 损失-似然自助法

一种颇具吸引力的目标参数确定方法是使用损失最小化。假设参数 θ 的真实值 $\theta_0 \in \Theta \subset \mathrm{R}^p$ 可使下列期望损失最小化，其中 l 是某个损失函数： $\mathcal{O} \times \Theta \to [0, \infty)$ ：

$$\theta_0 \equiv \theta(F_O) = \mathrm{argmin}_\theta \int_{\mathcal{O}} l(o;\theta) dF_0(o) = \mathrm{argmin}_\theta \, \mathbb{E}_{P_0} l(O;\theta).$$

若测度 P_O （其对应的分布函数为 F_O ）未知，但已知属于集合 $\mathcal{F}$，则 $\mathcal{F}$ 上的任意先验分布 $P_{\mathcal{F}}$ 都会在 Θ 上诱导出关于 P_Θ 的先验分布，具体做法是先考虑，对于 $F \in \mathcal{F}$,

$$\theta(F) = \mathrm{argmin}_t \int_{\mathcal{X}} l(o;t) dF(o)$$

然后定义 $P_\Theta(\theta \in A) = P_{\mathcal{F}}\{F\text{: } \theta(F) \in A\}$。在对 l 和先验分布 $P_{\mathcal{F}}$ 的温和条件下，该先验分布会在 θ_0 的邻域内赋予足够的概率质量。

Lyddon 等人（2019）采用损失最小化方法来确定目标参数，并使用贝叶斯非参数计算策略从 $\theta(P_O)$ 的后验分布中抽取样本。设 α 是 $(\mathcal{O}, B)$ 上的有限博雷尔测度。若在 F 上赋予狄利克雷过程先验 DP(α)（见 Ferguson（1973）），则给定 $O_1, \ldots, O_n$ 后， F 的后验分布将服从狄利克雷过程 DP($\alpha + n\mathrm{P}_n$)。当 $\alpha(\mathcal{O}) \to 0$ 时，后验分布的弱极限 DP($n\mathrm{P}_n$) 可表示为经验过程 P_n^w，其中 $(w_{1n}, \ldots, w_{nn}) \sim$狄利克雷分布(1, . . . , 1)；这一弱极限称为贝叶斯自助法分布，见 Rubin（1981）。还值得一提的是，后验分布 DP($\alpha + n\mathrm{P}_n$) 可写为 $V_n Q + (1 - V_n)\mathrm{P}_n^w$，其中 $V_n \sim$ 服从 Beta($|\alpha|$, n)， $Q \sim$ 服从 DP(α)，且 $|\alpha| = \alpha(\mathcal{O})$ 与 V_n、 Q 和 w 相互独立，见 Ghosal 和 van der Vaart（2017）。当 n 较大时，P_n^w 成为主导项，这一事实正是非参数伯恩斯坦-冯·米泽斯定理的核心。结合狄利克雷过程的这些性质，可将 $\theta(F)$ 的后验分布近似为

$$\mathrm{argmin}_\theta \int_{\mathcal{X}} l(O;\theta)\, d\mathbb{P}_n^w = \mathrm{argmin}_\theta \sum_{i=1}^{n} w_{in} l(O_i;\theta).$$

后验计算可以通过以下算法实现：

算法 2.1。 损失似然自助法（*Lyddon* 等人，*2019*）

1. *For* $j \in \{1, 2, \ldots, B\}$

(a) Draw random weights $(w_{1n}^{(j)}, \ldots, w_{nn}^{(j)}) \sim$ *Dir*$(1, \ldots, 1)$.

(b) Compute $\theta^{(j)} = arg\,min_\theta \sum_{i=1}^{n} w_{in}^{(j)} l(O_i; \theta)$.

2. 输出 $\theta^{(1)} \ldots, \theta^{(B)}$。

Lyddon 等人（2019）的以下定理推广了 Newton 和 Raftery（1994）的结果，并讨论了给定数据条件下，自助样本的渐近分布；该自助样本对应于使 $\mathrm{E}_{P_O} l(O; \theta)$ 最小化的 θ_0，其中 O 服从 P_O 分布。我们在假设 1 中给出相关概率空间所需的假设（见补充材料）。本文其余部分中，权重 $(w_{1n}, \ldots, w_{nn})$ 服从 Dirichlet$(1, \ldots, 1)$ 分布，记为 P_W，且生成时与数据独立。因此，联合概率测度为 $P_{OW} = P_O^\infty \times P_W$。必要时，为简单起见，我们用 P_O 代替 P_O^∞。

定理 2.1。设 $\hat{\theta}_n$ 和 $\theta_{n,BB}$ 分别为下列函数的极小化者

$$\frac{1}{n}\sum_{i=1}^{n} l(O_i; \theta) \qquad \textit{and} \qquad \sum_{i=1}^{n} w_{in} l(O_i; \theta),$$

其中 $(w_{1n}, \ldots, w_{nn})$ 是一个 *Dirichlet*$(1, \ldots, 1)$ 随机向量。那么，在正则条件下，以下量

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \qquad \textit{and} \qquad \sqrt{n}(\hat{\theta}_{n,BB} - \hat{\theta}_n) \mid O_1, \ldots O_n$$

依分布收敛为一个分布为 $n \to \infty$ 的随机变量

$$\mathcal{N}\left(0, J(\theta_0)^{-1} I(\theta_0) J\left(\theta_0\right)^{-1}\right)$$

几乎必然地 P_O^∞，其中

$$J(\theta_0) = \mathbb{E}_{P_O}\left\{\frac{\partial^2 l(O;\theta_0)}{\partial\theta\partial\theta^\top}\right\} \quad and \quad I(\theta_0) = \mathbb{E}_{P_O}\left\{\frac{\partial l(O;\theta_0)}{\partial\theta}\frac{\partial l(O;\theta_0)}{\partial\theta^\top}\right\}.$$

其中

$$\frac{\partial l(o;\theta_0)}{\partial\theta} = \left.\frac{\partial l(o;\theta)}{\partial\theta}\right|_{\theta=\theta_0}$$

依此类推。

在这种情况下，定理 2.1 给出的条件渐近分布与 $\hat{}\sqrt{n}(\theta_n - \theta_0)$ 的极限分布相同。这称为贝叶斯自助法相合性。然而，必须强调这两种极限陈述之间的差异。第一种情形中的随机性来自随机变量 $O_1 \ldots$， O_n，而在第二种情形中，这些变量被视为固定，唯一的随机性来源是权重。这种贝叶斯/频率学派二元性对于解读贝叶斯可信区域以及评估后验分布的渐近频率学派 $\hat{}$ 性质至关重要。围绕后验均值 θ_n 的贝叶斯可信区间，等同于基于 θ_n的频率学派 $\hat{}$ 置信区间。本质上，后验分布围绕其均值的条件行为将模拟后验均值围绕参数真实值的频率学派行为。此外，贝叶斯置信集将具有（渐近地）正确的覆盖概率。在贝叶斯分析中，参数型伯恩斯坦–冯·米泽斯定理确立了最大后验均值与最大似然估计量之间的这种联系，参见 Ghosh 和 Ramamoorthi（2003）。这种二元性也在半参数 M 估计量中得到研究，并已知成立；参见 Cheng 和 Huang（2010）以及 Ma 和 Kosorok（2005）。

2.2 多参数情形下的损失似然自助法

当 θ 是一个多分量向量时，我们可以通过通常的边缘化方式，考察 θ 某个子分量的后验分布。具体而言，可通过提取联合协方差矩阵 $J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}$中相关的分块，计算边缘后验方差（渐近方差）。类似地，利用多元正态分布的通常性质，可以计算一个子分量在给定另一子分量条件下的条件分布方差（渐近方差）。在这种情况下，根据标准论证，显然边缘后验方差始终不小于条件后验方差，即前者与后者之差是半正定的。

损失似然自助法明确地将 θ的所有元素作为整体进行（加权）最小化。也就是说，θ_0 是推断的明确目标，而损失函数则同时概括了观测数据与 θ 所有元素之间的关系。然而，由于联合分布是有效的概率分布，由此得到的边缘分布也都是有效分布。因此，单独解释计算得到的边缘后验并不直接；这在一定程度上是因为缺少用于可观测量的生成概率模型。

如果 $\theta^\top = (\theta_1^\top, \theta_2^\top)$，该表述

$$\theta_{10}(F) = \text{argmin}_{\theta_1} \int_{\mathcal{X}} l(o; \theta_1, \theta_{20}) dF(o)$$

很自然，但还需要进一步定义 θ_{20}。从根本上说，这两个参数通过损失函数内在地联系在一起，而通过某种边际化方法确定目标参数并非易事。也可能出现这样的情况：θ_{10} 无法通过一个唯一确定的损失函数来识别；并且，用于识别同一参数的两个不同损失函数，可能不会得出相同的边际推断。具体而言，仍有可能实现一致估计，但得到的（渐近）方差可能不同。

如果损失函数是可分离的，例如

$$l(o; \theta_1, \theta_2) = l_1(o; \theta_1) + l_2(o; \theta_2)$$

那么，对 θ_1 和 θ_2 进行边际解释或许是可行的。在传统参数设定下，若满足可交换性并采用层级模型的德·芬尼蒂表示，可分离性便会出现，例如

$$f_1(y|x; \theta_1) f_2(x; \theta_2)$$

其中 θ_1 和 θ_2 分别可在条件模型和边际模型中作出解释。如果不存在可分离性，则可以采用剖面化方法；如果目标是 θ_1，那么对于每个 θ_1，我们考虑针对 θ_2进行最小化，并考察剖面损失函数随 θ_1 变化的情况。不过，这仍留下一个问题：狄利克雷权重应在何处以及如何使用。我们将在以下章节尝试解决这一问题。

3 估计方程的自助法

损失似然自助法侧重于最小化损失函数，但它与基于求根的方法之间显然存在联系，例如基于估计方程的频率学派方法。本节将上述论述应用于定义为估计方程解的参数。随后，我们讨论涉及有限维干扰参数的估计方程的扩展，并探讨如何从目标参数的后验分布中获取样本，使其具有良好的频率学性质。

3.1 一个示例模型

设 $O = (Y, X, Z, U_1, U_2)$，并设 $O_1, \ldots\ldots, O_n$ 为独立同分布的随机变量。假设数据生成机制为

$$Y = \theta_0 Z + g_0(X) + U_1, \tag{1}$$
$$Z = e_0(X) + U_2, \tag{2}$$

其中 $\theta_0 \in \mathbb{R}$、g_0 和 e_0 是 $X \in \mathbb{R}^q$、$U_1 \sim N(0, \sigma^2)$ 的实值函数，且 $\mathbb{E}[U_2|X] = 0$。（1）中的模型称为部分线性模型。在对其形式仅作最少假设的情况下，灵活地估计 g_0 通常并非易事。

假设 Z是一个伯努利随机变量，且 $P(Z = 1|X) = e_0(X)$；在这种情况下，$U_2 = \mathbb{1}\{U < e_0(X)\} - e_0(X)$，其中 $U \sim Uniform(0, 1)$ 与 U_1和 X相互独立，因此 U_2与 $e_0(X)$不相关，且其方差为$e_0(X)(1-e_0(X))$。在因果推断设定中，函数 e_0 称为倾向评分，而 θ_0称为平均处理效应（ATE），代表目标参数。进一步假设 $e_0(x) = e(x; h_0) = \text{expit}(h_0^\top x)$，其中 $\text{expit}(t) = e^t/(1 + e^t)$，且 h_0 是 $\mathbb{R}^q$中的干扰参数。定义

$$m(O; \theta_0, h_0) = (Y - \theta_0 Z) \{Z - e(X; h_0)\}, \tag{3}$$

Robins 等人（1992）利用 $\mathbb{E}m(O; \theta_0, h_0) = 0$ 这一事实，来获得 θ_0 的估计量。由于实际中未知 h_0，因此通过对 Z 关于 X 进行逻辑回归来估计 $\hat{}$，由此得到 h_0 的估计量 h。随后，将依赖于 h 的 $\hat{}\theta_0$ 的估计量 θ_n 作为以下方程的解：

$$\sum_{i=1}^{n} (Y_i - \theta Z_i) \left\{ Z_i - e(X_i; \hat{h}) \right\} = 0,$$

使得

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n} Y_i \left\{ Z_i - e(X_i; \hat{h}) \right\}}{\sum_{i=1}^{n} Z_i \left\{ Z_i - e(X_i; \hat{h}) \right\}}. \tag{4}$$

还有其他可能值得关注的误设模型。例如，我们可以考虑矩约束 Ek(*O*; θ_0, ϕ_0, h_0) $= 0$，其中

$$k(O; \theta, \phi, h) = \{Y - \theta Z - \phi e(X; h)\} \{Z - e(X; h)\}, \tag{5}$$

以及 $\phi_0 = \mathrm{E}\{g_0(X)e_0(X)\} / \mathrm{E}\left\{e_0(X)^2\right\}$。求解出以 h 表示的 ψ 和 ϕ，等价于对结果 Y 关于处理 Z 和倾向评分进行回归。这种方法试图通过将 $g_0(X)$ 投影到 $e_0(X)$ 的张成空间来加以考虑，而此前使用的函数 m 并未考虑 $g_0(X)$。关于由 k 描述的误设模型的详细信息，请参阅 Moodie和Stephens（2022）；该模型称为倾向得分回归模型。请注意，在某些条件下，(3) 和 (5) 中的函数会产生一致且渐近正态的（频率学）估计量。

还有其他仅依赖倾向评分知识的处理效应估计量。θ_0的逆概率加权估计量（IPW estimator）就是其中之一。该估计量源自以下矩约束

$$\mathbb{E}\left[\theta_0 - \left\{ \frac{ZY}{e(X; h_0)} - \frac{(1-Z)Y}{1 - e(X; h_0)} \right\}\right] = 0. \tag{6}$$

另一种此类估计量是处理组平均处理效应（ATT）的估计量。ATT，ρ_0，定义为$\rho_0 = \mathrm{E}\{Y(1) - Y(0)|Z = 1\}$，满足矩约束 E$r$(*O*; ρ_0, h_0) $= 0$，其中

$$r(O; \rho, h) = ZY - \frac{Y(1-Z)e(X; h)}{1 - e(X; h)} - \rho Z.$$

关于ATT的概述，可参见Moodie等人（2018）。所得估计量不受结果模型形式的影响，只取决于倾向模型是否正确设定。这种矩约束不能视为对似然的完整设定，而应视为对可观测变量分布的约束，该分布由因果推断框架中的一般假设所决定。

3.2 使用估计方程进行参数靶向

令 m： $\mathcal{O} \times \Theta \to \mathrm{R}^p$。对于 $F \in \mathcal{F}$，令 θ(*F*)表示以下方程的解

$$\int_{\mathcal{O}} m(o; \theta) dF(o) = 0.$$

并令 $\theta_0 \equiv \theta(F_O)$。采用第 3 节中讨论的相同论证，我们可以通过实施以下算法，从 θ 的后验分布中获得一个样本：

算法 3.1. 用于估计方程的贝叶斯自助法

1. 对每个 $j \in \{1, \ldots, B\}$,

(a) Draw random weights $(w_{1n}^{(j)}, \ldots, w_{nn}^{(j)}) \sim$ *Dir*$(1, \ldots, 1)$.

(b) Find $\theta^{(j)}$ *as the solution to* $\sum_{i=1}^{n} w_{in}^{(j)} m(O_i; \theta) = 0$

2. 输出 $\theta^{(1)}, \ldots, \theta^{(B)}$。

以下标准结果（例如见 Kosorok (2008)）确立了算法 3.1 所需的贝叶斯/频率学派二元性。该定理也可以采用与定理 4.1 和定理 4.2 中将使用的相同技术和相近正则性条件来证明，因此我们将讨论推迟到后文。

定理 3.1. 设 θ_n 与$\theta_{n,\ BB}$ 满足

$$\sum_{i=1}^{n} m(O_i; \hat{\theta}_n) = 0 \qquad \textit{and} \qquad \sum_{i=1}^{n} w_{in} m(O_i; \hat{\theta}_{n,BB}) = 0$$

分别。则在正则条件下，

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \qquad \textit{and} \qquad \sqrt{n}(\hat{\theta}_{n,BB} - \hat{\theta}_n)|O_1, \ldots O_n$$

依分布收敛为一个具有 $\mathcal{N}(0, L)$ 分布的随机变量，其中 *e*

$$L = \left[\mathbb{E}_{P_O}\left\{\frac{\partial m(O;\theta_0)}{\partial \theta^\top}\right\}\right]^{-1} \mathbb{E}_{P_O}\left\{m(O;\theta_0)m(O;\theta_0)^\top\right\}\left[\mathbb{E}_{P_O}\left\{\frac{\partial m(O;\theta_0)}{\partial \theta^\top}\right\}\right]^{-\top}$$

第二个结果给出了目标参数的完全贝叶斯（渐近）后验分布。如果定理 2.1 中的 l 关于 θ 可微，那么该最小化问题等价于定理 3.1 中讨论的问题。不过，采用估计方程的视角可以涵盖更一般的情形。若 m 属于Donsker函数类，则可将 $m(O; \theta)$ 的可微性条件替换为 $\mathrm{E}_{P_0} m(O;\theta)$ 的可微性，从而可以利用 Donsker类的可交换自助法性质；参见 van der Vaart 和 Wellner（1996年）。

3.3 存在干扰参数时的目标参数

我们现在假设存 在有限维干扰参数 h_0，并假设 θ_0 是以下方程的解

$$\int_{\mathcal{O}} m(o;\theta,h_0)dF_O(o) = 0.$$

在多参数设定下，这是一种便于对 θ_0进行目标化的自然方式。与先前讨论的方法类似，我们可以通过抽取狄利克雷权重来获得后验分布。然而在实践中， h_0 是˜ ˜未知的，需要进行估计。如果 h 是 h_0的估计量，那么在从目标参数的后验分布中获取样本时，处理 h 不确定性的确切方法便是一个至关重要的问题。基于第2.1小节中所述的原因，我们寻求一个满足频率学派/贝叶斯二元性定理的后验分布。

˜在处理将 h 纳入 θ的后验推断这一问题时，可以采用多种可能的方法，˜并且会引出若干问题。首先，是否应在单独的第一步中获得 h ，并在整个后验分布计算过程中保持固定？换言之，利用贝叶斯自助法得到的 θ 的后验分布是以对 h的˜知识为条件的，而该知识在贝叶斯程序中保持不变。此处并未尝试传播附着于 h的˜不确定性。由此得到的目标参数后验分布具有哪些性质，以及
˜ ˜此后验分布如何依赖于获得 h 的方式？其次，如果我们要考虑附着于 h的不确定性，应当如何处理它，才能得到具有良好频率学派性质的后验分布？

我们在第4节探讨这些问题。我们重点关注贝叶斯/频率学派二元性，因此研究具有渐近性质。不过，我们也研究有限样本层面的影响。值得注意的是，基于不同设定的估计方程的贝叶斯方法，可能会产生方差不同的后验分布。例如，在部分线性模型中，尝试使用灵活的参数族来表示 g_0 ，通常会得到比估计量 (4) 所计算出的后验方差更小的后验方差。这类结果在频率学派文献中已有深入研究，我们的结果使我们能够在贝叶斯框架下得出类似结论。

4 渐近结果

回到第3.1节的示例模型，方程 (4) 中的估计量具有易于处理的性质。估计方程的频率学派理论使我们能够得出结论： $\sqrt{n}(\theta_n - \theta_0)$ 渐近服从正态分布，其渐近方差为 V，参见 Tsiatis (2006)。Robins 等人（1992）指出， V 小于估计 θ_0 的估计量的渐近方差 Σ，若假设 h_0 已知而无需估计。这个悖论现象源于这样一个事实：增加 h_0 的变异性会降低 θ_0 的变异性。Henmi and Eguchi (2004) 将这一现象推广，并证明估计干扰参数可能会改善目标参数估计量的渐近方差。

在接下来的章节中，为了在存在干扰参数 h_0的情况下对目标参数 θ_0 进行贝叶斯推断，我们将偏离这一频率学派框架。

4.1 联结贝叶斯自助法

在频率学派框架中，基于倾向评分的调整通常通过回归、匹配、逆加权或分层来进行。两步法中，使用估计的倾向评分来

推断处理效应的方法得到了广泛研究。在贝叶斯框架中，两步法有悖于“似然乘以先验分布”的传统贝叶斯框架。关于现有参数方法以及贝叶斯因果推断所面临挑战的详细综述，参见 Li 等人（2023）。假设处理效应与倾向评分具有联合参数模型的方法（参见 McCandless 等人（2009））会产生反馈，从而导致有偏估计。其他方法则通过切断反馈法来缓解这种反馈（参见 McCandless 等人（2010））；还有一些两步法在已知倾向评分估计值的条件下进行贝叶斯推断（参见 Hoshino（2008）以及 Kaplan 和 Chen（2012））。尽管对于两步法和切断反馈法是否可以被视为完全贝叶斯方法尚未达成共识，Stephens 等人（2023）偏离了贝叶斯参数框架，转而采用贝叶斯非参数方法来推断有限维参数。他们认为，在谨慎获得倾向评分的贝叶斯估计值的前提下，两步法是因果结构模型中的有效贝叶斯方法。

对于第 3.1 小节中由估计函数 (5) 给出的倾向得分回归模型，可以通过以下过程获得 $(\theta,\ \phi)$ 的后验样本，该过程称为关联贝叶斯自助法：

算法 4.1. 关联贝叶斯自助法

1. 对于每个 $j \in \{1, \ldots, B\}$,

   (a) 抽取随机权重 $(w_{1n}^{(j)}, \ldots, w_{nn}^{(j)}) \sim$ 狄利克雷分布$(1, \ldots, 1)$。

   (b) 令 $h^{(j)} = argmin_h \sum_{i=1}^{n} w_{in}$ 对数 $f_2(z_i|x_i; h)$

   (c) 定义$(\theta^{(j)}, \phi^{(j)})$ 为

$$(\theta^{(j)}, \phi^{(j)}) = argmin_{(\theta,\phi)} \sum_{i=1}^{n} w_{in}^{(j)} \log f_1(y|x, z, \theta, \phi, \hat{h}^{(j)})$$

2. 输出$(\theta^{(1)}, \phi^{(1)}) \ldots, (\theta^{(B)}, \phi^{(B)})$。

其中，$f_2(z|x)$ 是 $Z|X$ 的正确设定条件密度，而 $f_1(y|x, z)$ 是 $Y|X, Z$ 的有意错误设定密度。“关联”一词强调，在两个优化步骤中均保留同一组狄利克雷权重，从而能够正确传播 h 的推断不确定性，以获得目标参数的后验分布（θ, ϕ）。

在本节余下部分，我们将从其他角度考察关联贝叶斯自助法。在理论评估其良好的频率学派性质后，我们将从估计方程而非损失最小化的角度来诠释关联贝叶斯自助法，从而将其视为一种一步法，同时能够对 θ 和 h 的联合后验分布进行抽样。

4.2 传统插件方法

本节考察在ˆ两步法中不传递附着于干扰参数估计量的不确定性所产生的影响。例如，假设 h 是 h_0 的一个估计量，并且是渐近线性的，即

$$\sqrt{n}(\hat{h} - h_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(O_i; h_0) + o_{P_O}(1),$$

其中，$\mathrm{E}_{P_O}\psi(O;\ h_0) = 0$，且 $\mathrm{Var}_{P_0}\psi(O;\ h_0) < \infty$。这种传统代入法背后的核心思想，是使用对干扰参数的精确估计来消除反馈，参见 Zigler 等人（2013）。Saarela 等人（2015）指出，在“似然乘以先验”的传统贝叶斯方法背景下，频率学派两步程序通常依赖基于倾向评分的校正，但不存在相应的贝叶斯方法，因而不能被视为完全贝叶斯方法。Graham 等人（2016）提出了方差校正，以便在双重稳健模型的两步程序中纳入倾向评分的推断不确定性。

ˆ 我们表明，在估计 θ 时若未考虑与 h 相关的推断不确定性，无论使用何种估计量 h，通过贝叶斯自助法得到的后验分布估计都会表现出不理想的特性。正如第 3.1 小节所述，经典频率学派参数结果促使人们使用估计得到的干扰参数，以降低目标参数的变异性；然而，我们的贝叶斯计算表明，固定估计值不会影响渐近后验方差。此外，我们还得出了一个令人惊讶的结果：渐近后验方差不取决于 h，而 h 仅影响后验均值。更具体地说，此处的渐近后验方差

等于始终假设 h_0 已知时的渐近后验方差。这一结果看似有利，但我们将在陈述定理 4.1 后的讨论中进一步阐述其弊端。

ˆ ˆ 可以考虑几种不同的估计量 h：例如，若已知 h_0，则可以取 $h = h_0$。若未知ˆh_0，则可以通过对处理关于协变量进行常规频率学派逻辑回归来获得 h。从贝叶斯角度来看，对干扰参数进行贝叶斯分析可能更为恰当。仍可考虑具有贝叶斯性质的 h_0估计量：首先，通过损失似然自助法等方法从 h的后验分布中抽取样本，然后令 h 为所得样本的均值。可以证明，这三种建议的估计量具有不同的影响函数和不同的方差，但在未考虑h_0 点估计的推断不确定性的情况下，目标参数的渐近后验方差仍保持不变。

在第 3.1 节的示例模型中，可以采用以下计算方法：

算法 4.2。 传统插件法的贝叶斯自助法

ˆ*1.* 获得 h_0 的一个渐近线性估计量 h。

2. 对于每个 $j \in \{1, \ldots, B\}$,

(a) 抽取随机权重$(w_{1n}^{(j)}, \ldots, w_{nn}^{(j)}) \sim$狄利克雷分布$(1, \ldots, 1)$。*(b)* 求 $\theta^{(j)}$ 作为其解

$$\sum_{i=1}^{n} w_{in}^{(j)} \{Y_i - \theta Z_i\} \left\{ Z_i - e(X_i; \hat{h}) \right\} = 0$$

3. 输出 $\theta^{(1)}$，……，$\theta^{(B)}$。

ˆ为计算后验方差，我们借助经验过程的基本工具。我们将引入 θ_n 作为下列估计方程的解

$$\sum_{i=1}^{n} m(O_i, \theta, \hat{h}) = 0.$$

ˆ ˆ在实践中，θ_n 是利用 h_0的代入估计量 h 得到的 θ_0 的频率学派估计量。尽管为了得到ˆθ_0的频率学派估计量而在第一步使用 h_0 的贝叶斯估计量可能有悖直觉或不寻常，但我们强调，引入 θ_n 仅仅是作为一种便于识别渐近后验分布的理论工具。我们将在陈述定理 4.1 之后详细阐述这一点。下述定理将上述讨论的结果形式化：

ˆ**定理 4.1.** 假设 h 是满足以下条件的 h_0 估计量

$$\sqrt{n}(\hat{h} - h_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(O_i; h_0) + o_{P_O}(1),$$

ˆ ˆ其中 $\mathbb{E}_{P_O} \psi(O;\ h_0) = 0$，且 $Var_{P_O} \psi(O;\ h_0) < \infty$。设 θ_n 和 $\theta_{n,\ BB}$分别满足以下条件

$$\sum_{i=1}^{n} m(O_i, \hat{\theta}_n, \hat{h}) = 0 \qquad \textit{and} \qquad \sum_{i=1}^{n} w_{in} m(O_i, \hat{\theta}_{n,BB}, \hat{h}) = 0,$$

分别。设

$$M_{\theta_0} = \mathbb{E}_{P_O} \left\{ \frac{\partial m(O; \theta_0, h_0)}{\partial \theta^\top} \right\} \equiv \mathbb{E}_{P_O} \left\{ \left. \frac{\partial m(O; \theta, h_0)}{\partial \theta^\top} \right|_{\theta = \theta_0} \right\}.$$

那么，在假设*2*所述的正则条件下，下列结论成立：

(a)

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right) = \frac{-1}{\sqrt{n}} M_{\theta_0}^{-1} \sum_{i=1}^{n} \{ m(O_i; \theta_0, h_0) + M_{h_0} \psi(O_i; h_0) \} + o_{P_O}(1),$$

(b)

$$\sqrt{n}\left(\hat{\theta}_{n,BB}-\hat{\theta}_n\right)=-\sqrt{n}M_{\theta_0}^{-1}\sum_{i=1}^{n}\left(w_{in}-\frac{1}{n}\right)m(O_i;\theta_0,h_0)+o_{P_{OW}}(1)$$

引入 ˆ θ_n 使我们能够将 $\sqrt{n}$(ˆ$\theta_{n,BB}$ − ˆθ_n) 用经验过程 $\sqrt{n}$(P$_n^w$ − P$_n$) 表示，从而迅速识别渐近后验分布。事实上，这一简便表达式使我们得出 ˆ ˆ 结论：$\sqrt{n}$($\theta_{n,BB}-\theta_n$) 在给定 $O_1,\ldots,O_n$ 时的条件分布渐近正态，具有 ˆ 零均值和方差 Σ。当 n 较大时，对于给定实现 $O_1,\ldots,O_n$，后验均值近似为 θ_n，ˆ其依赖于 h。渐近后验方差 Σ 并不取决于 h_0 的估计方式，也不会 ˆ 自动匹配 V，即 $\sqrt{n}$($\theta_n-\theta_0$) 的频率学派渐近方差。类似于定理 2.1 中的贝叶斯/频率学派二元性并不必然成立。当然，若假定 h_0 已知，则该二元性成立。尽管这些方差的表达式可能看似复杂，比较它们也可能较为困难，但参考 Henmi and Eguchi (2004) 无疑会简化这一任务。为比较这两个方差，我们将 Σ 解释为频率学派方差。具体而言， Σ 是在干扰参数 ˆh_0 已知时 θ_0 的估计量的频率学派渐近方差，而 V 则是在第一步中估计了 h 时 θ_0 的频率学派估计量的渐近方差。Henmi ˆ and Eguchi (2004) 使我们能够推断 Σ ≥ V，只要 h 是通过不依赖于 θ 的估计方程获得的。在这种情况下，常规两步法会诱导出一个渐近正态的后验分布，其均值正确，但方差更大。除非假定 h_0 已知，否则贝叶斯可信区间的覆盖率将高于名义覆盖率。

4.3 关联贝叶斯自助法：贝叶斯与频率学的二元性

ˆ ˆ 我们已经看到，两步法中的第一步是用 h 估计 h_0，然后使用 h；如果贝叶斯分析是在给定 h 的 ˆ 值的条件下进行，就会导致目标参数的后验方差膨胀。在本文接下来的部分中，我们将针对这一问题提出解决方案，并论证此类方法完全符合贝叶斯框架。由于我们将讨论范围限定在对 h_0 的参数估计上，因此可以合理地假设 h_0 满足 Eu(O; h_0) = 0，其中 u 是某个函数：$\mathcal{O}\times\mathcal{H}\to$ R^q。在第 3.1 小节的示例中，我们可以立即看出，h_0 是 Eu(O; h) = 0 的解，其中

$$u(O;h)=X^\top\{Z-e(X;h)\}.$$

ˆ 在频率学派框架下，如果 h_n 满足

$$\sum_{i=1}^{n}u(O_i;\hat{h}_n)=0,$$

那么，在关于 u 的正则条件下，我们可以得到 $\sqrt{n}$(ˆh_n − h_0) 依分布收敛于一个具有 $\mathcal{N}$(0, Ω) 分布的随机变量，其中

$$\Omega=U_{h_0}^{-1}\mathbb{E}_{P_0}\left\{u(O;h_0)u(O;h_0)^\top\right\}U_{h_0}^{-\top}\quad\text{and}\quad U_{h_0}=\mathbb{E}_{P_O}\left\{\frac{\partial u(O;h_0)}{\partial h^\top}\right\}.$$

ˆ 接下来，我们考虑满足以下条件的频率学派估计量 θ_n：

$$\sum_{i=1}^{n}m(O_i,\hat{\theta}_n,\hat{h}_n)=0.$$

ˆ ˆ 虽然以下关于 θ_n 和 h_n 渐近性质的结论并非证明定理 4.2 所必需，但我们在此列出，因为它们对后续讨论至关重要。利用 Henmi and Eguchi (2004) 的结果可知，$\sqrt{n}$(ˆ$\theta_n-\theta_0$) 收敛于一个服从 $\mathcal{N}$(0, V) 分布的随机变量，其中

$$V=\Sigma-M_{\theta_0}^{-1}M_{h_0}\Omega M_{h_0}^\top M_{\theta_0}^{-\top},$$

并且 $\sqrt{n}$(ˆh_n-h_0)与 $\sqrt{n}$(ˆ$\theta_n-\theta_0$)渐近独立。

关联贝叶斯自助法依赖这一简单的频率学机制，来处理贝叶斯估计目标参数时附着于干扰参数的不确定性。当然，采用该算法的根本动机更加深刻，并借鉴了因果推断、贝叶斯非参数方法和决策理论中的概念。Stephens 等人（2023）旨在论证，适当选择的倾向评分代入估计量的使用

作为一种贝叶斯方法是合理的。在第4.5小节中，在陈述定理 4.2并讨论其含义之后，我们将介绍一种不涉及插件法概念、完全依赖贝叶斯非参数方法和估计方程的该算法动机。在进行理论陈述之前，我们展示第 3.1 小节模型的一个关联贝叶斯自助法算法实例。

算法 4.3. 部分线性模型的关联贝叶斯自助法

1. 对于每个 $j \in \{1, \ldots, B\}$,

(a) 抽取随机权重 $(w_{1n}^{(j)}, \ldots, w_{nn}^{(j)}) \sim$ 狄利克雷分布$(1, \ldots, 1)$。

(b) 求得 $h^{(j)}$ 作为以下方程的解

$$\sum_{i=1}^{n} w_{in}^{(j)} X_i^{\top} \{Z_i - e(X_i; h)\} = 0$$

(c) 求解方程，得到 $\theta^{(j)}$

$$\sum_{i=1}^{n} w_{in}^{(j)} \{Y_i - \theta Z_i\} \left\{ Z_i - e(X_i; h^{(j)}) \right\} = 0$$

2. 输出 $\theta^{(1)}, \ldots, \theta^{(B)}$。

关于 $h^{(j)}$ 的不确定性通过用于获得 $\theta^{(j)}$ 的同一组狄利克雷权重来处理，这些权重在干扰参数与目标参数之间建立了关联。现在我们陈述以下定理，用以评估上述算法的渐近性质。

定理 4.2. 假设以下条件成立：

1. $\sum_{i=1}^{n} u(O_i; \hat{h}_n) = 0$ *and* $\sum_{i=1}^{n} m(O_i; \hat{\theta}_n, \hat{h}_n) = 0$
2. $\sum_{i=1}^{n} w_{in} u(O_i; \hat{h}_n) = 0$ *and* $\sum_{i=1}^{n} w_{in} m(O_i, \hat{\theta}_{n,BB}, \hat{h}_{n,BB}) = 0.$

那么，在假设 *3* 中陈述的正则性假设下，

(a)

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{-1}{\sqrt{n}} M_{\theta_0}^{-1} \sum_{i=1}^{n} \left\{ m(O_i; \theta_0, h_0) - M_{h_0} U_{h_0}^{-1} u(O_i; h_0) \right\} + o_{P_O}(1),$$

(b)

$$\sqrt{n}\left(\hat{\theta}_{n,BB} - \hat{\theta}_n\right)$$

$$= -M_{\theta_0}^{-1} \sqrt{n} \sum_{i=1}^{n} \left(w_{in} - \frac{1}{n} \right) \left\{ m\left(O_i; \theta_0, h_0\right) - M_{h_0} U_{h_0}^{-1} u\left(O_i; h_0\right) \right\} + o_{P_{OW}}(1)$$

这些结果共同意味着

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \qquad \textit{and} \qquad \sqrt{n}(\hat{\theta}_{n,BB} - \hat{\theta}_n) \mid O_1, \ldots, O_n$$

依分布收敛到具有 $\mathcal{N}(0,\ V)$ 分布的随机变量（几乎必然）。

这一结果通过确立有利的贝叶斯/频率学派二元性，证明了关联贝叶斯自助法具有理想的频率学派性质。我们观察到，使用关联贝叶斯自助法得到的 θ 的后验方差，不大于假设 h_0 始终已知时的后验方差。

4.4 近似精度

这很 研究标准化后验分布是值得的，也就是.....的条件分布

$$\sqrt{n} V^{-0.5} (\hat{\theta}_{n,BB} - \hat{\theta}_n)$$

给定 $O_1, \ldots\ldots O_n$，作为标准化量的频率学派抽样分布的近似

$$\sqrt{n}V^{-0.5}(\hat{\theta}_n - \theta_0),$$

通过 Edgeworth展开。我们假设，正如 Newton 和 Raftery（1994）以及 Shao和Dongsheng（1995）所述，标准化的贝叶斯估计量和频率学派估计量都可以通过各自的影响函数得到良好近似。假设 θ_0 是一个实值参数，令 Φ 和 ϕ 分别表示标准正态分布和密度，并遵循 Newton 和 Raftery（1994）中的相同思路，则在 Hall（1992）和 Shao和Dongsheng（1995）中所述正则条件下，频率学分布 F_n满足

$$\sup_{t\in\mathbb{R}}\left|F_n(t) - \Phi(t) - \frac{1}{6\sqrt{n}}\left[\frac{\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^3\right\}}{[\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^2\right\}]^{3/2}}\right](t^2-1)\phi(t)\right| = o(n^{-1/2}),$$

而后验分布$F_{n,\ BB}$ 满足

$$\sup_{t\in\mathbb{R}}\left|F_{n,BB}\left(t \mid O_1,\ldots O_n\right) - \Phi(t) - \frac{2}{6\sqrt{n}}\left[\frac{\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^3\right\}}{[\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^2\right\}]^{3/2}}\right](t^2-1)\phi(t)\right|,$$

几乎必然为 $o(n^{-1/2})$（见 Haeusler 等人，1991），其中

$$\lambda(O;\theta_0,h_0) = m(O;\theta_0,h_0) - M_{h_0}U_{h_0}^{-1}u(O;h_0)$$

这表明，$F_{n,\ BB}$ 对 F_n 的近似程度与 Φ 对 F_n的近似程度相同，均达到二阶精度，因为根据三角不等式，我们有

$$\sup_{t\in\mathbb{R}}\left|F_{n,BB}\left(t \mid O_1,\ldots O_n\right) - F_n(t) - \frac{1}{6\sqrt{n}}\left[\frac{\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^3\right\}}{[\mathbb{E}_{P_O}\left\{\lambda(O;\theta_0,h_0)^2\right\}]^{3/2}}\right](t^2-1)\phi(t)\right|$$

几乎肯定是 $o(n^{-1/2})$成立。这一近似可以通过讨论 Newton 和 Raftery 研究结果的方式加以说明 y（1994），Albert Y. L o 引入了渐近精度系数，并将其定义为每个 $t \in$ R 处的

$$\lim_{n\to\infty}\frac{\sqrt{n}\left|F_n(t) - F_{n,BB}(t \mid O_1,\ldots O_n)\right|}{\sqrt{n}\left|F_n(t) - \Phi(t)\right|}.$$

并指出，对于任意 $t \in$ R，该系数几乎必然等于 1。这强调了将后验分布视为对 F_n 的近似时，它与对 F_n的正态近似相吻合。

然而，这类近似并非二阶精确，因为 $F_{n,BB}$ 和 F_n 的 Edgeworth展开中 $n^{-1/2}$ 的系数并不相等。关于 Edgeworth展开更严谨的处理（包括使此类近似有效所需的正则性假设），请参见 Shao和Dongsheng（1995）。

4.5 关联贝叶斯自助法的一步论证

第 4.3 小节所述的程序为目标参数给出了具有良好渐近性质的后验分布，其可信区间的覆盖率达到名义覆盖率。这与第 4.2 小节所述的方法有很大不同，后者没有考虑估计干扰参数时的不确定性。虽然仍可以构建并论证一些理由，以验证在不传递不确定性的情况下将插件法作为全贝叶斯方法使用，但鉴于定理 4.1 所揭示的后验分布具有不理想的性质，我们选择不考虑这类理由。不过，我们希望为第 4.3 小节讨论的关联贝叶斯自助法找到一种全贝叶斯解释，使其与第 3.2 小节一致，并避开代入估计量这一有问题的概念。这种解释具有一般性，并借鉴了估计方程和贝叶斯非参数方法中的基本思想。

沿用前几节的记号，令

$$a(O;\theta,h) = \begin{bmatrix} m(O;\theta,h) \\ u(O;h) \end{bmatrix} \in \mathbb{R}^{p+q}.$$

这一解释将 h_0 视为与 θ_0地位相同的目标参数。我们通过 $\eta = (\theta^\top, h^\top)^\top$定义 $\eta \in \mathrm{R}^{p+q}$，并用 $a(O;\ \eta)$代替 $a(O;\ \theta, h)$。注意，矩条件 $\mathrm{E}a(O;\eta)\ = 0$ 的解 η_0 由 $\eta_0 = (\theta_0^\top, h_0^\top)^\top$ 给出，且估计量 $\hat{}\eta_n$ 满足

$$\sum_{i=1}^{n} a(O_i, \hat{\eta}_n) = 0,$$

给出的是 n由$\hat{}\eta_n = (\hat{}\theta_n^\top, \hat{}h_n^\top)^\top$给出，其中$\hat{}\, h_n$与$\hat{}\, \theta_n$分别满足定理4的第一个假设 .2:

$$\sum_{i=1}^{n} u(O_i, \hat{h}_n) = 0 \text{ and } \sum_{i=1}^{n} m(O_i, \hat{\theta}_n, \hat{h}_n) = 0.$$

由此可以清楚地看出

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_n - \theta_0 \\ \hat{h}_n - h_0\end{pmatrix} = \sqrt{n}\,(\hat{\eta}_n - \eta_0) \to \mathcal{N}(0, \Lambda),$$

其中，矩阵 $(p+q) \times (p+q)$ Λ 由下式给出

$$\Lambda = \left[\mathbb{E}_{P_0}\left\{\frac{\partial a(O;\eta_0)}{\partial \eta^\top}\right\}\right]^{-1}\left[\mathbb{E}_{P_0}\left\{a(O;\eta_0)a(O;\eta_0)^\top\right\}\right]\left[\mathbb{E}_{P_0}\left\{\frac{\partial a(O;\eta_0)}{\partial \eta^\top}\right\}\right]^{-\top}$$

并且利用 Henmi and Eguchi (2004) 的结果可以证明其满足

$$\Lambda = \begin{pmatrix} V & 0_{p\times q} \\ 0_{q\times p} & \Omega \end{pmatrix}.$$

摆脱这一频率学派框架，并遵循第 3.2 小节中的相同推理，我们可以通过求解以下方程，使用贝叶斯自助法得到 η 的后验分布

$$\sum_{i=1}^{n} w_{in} a(O_i, \hat{\eta}_{n,BB}) = 0.$$

可以直接看出，$\hat{}\eta_{n,\ BB} = (\hat{}\theta_{n,\ BB}^\top,\ \hat{}h_{n,\ BB}^\top)^\top$，其中$\hat{}h_{n,\ BB}$和$\hat{}\theta_{n,\ BB}$分别满足定理 4.2 的第二个假设:

$$\sum_{i=1}^{n} w_{in} u(O_i, \hat{h}_{n,BB}) = 0 \quad \text{and} \quad \sum_{i=1}^{n} w_{in} m(O_i, \hat{\theta}_{n,BB}, \hat{h}_{n,BB}) = 0.$$

因此，利用第 3.2 小节，我们可以得出

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_{n,BB} - \hat{\theta}_n \\ \hat{h}_{n,BB} - \hat{h}_n\end{pmatrix} \mid O_1, \ldots, O_n = \sqrt{n}\,(\hat{\eta}_{n,BB} - \hat{\eta}_n) \mid O_1, \ldots, O_n$$

收敛于一个服从 $\mathcal{N}(0,\ \Lambda)$ 分布的随机变量。因此，θ和 h 在后验意义上渐近独立；对 θ取渐近边际后验分布，即可得到定理 4.2 的结果。证明（见补充材料）依赖于代入的干扰参数这一概念，以便在不考虑联合后验分布的情况下，对 θ 的边际后验分布进行推断。然而，上述过程首先确立了联合后验分布的性质，然后利用这些性质得到目标参数的边际后验分布。因此，关联贝叶斯自助法可被视为一种一步法，用于对 h 和 θ进行联合后验推断。两步法是一种简单的计算策略，用于求解由增广函数 $a(O;\eta)$表示的方程组；对于在 θ的后验推断中使用代入估计量，以及倾向模型的不确定性传播这些颇具争议的概念，它并不特别重视。这种对关联贝叶斯自助法的理解克服了将两个步骤合并为一步的困难：后者需要在没有反馈的情况下联合估计倾向评分和因果效应；参见 Zigler（2016）。这一理解源于非参数贝叶斯定理的应用。

将问题表述为估计方程，而不是像算法 4.1 那样进行两次优化，这一点至关重要。事实上，这两次优化可能无法表述为一个可应用损失似然自助法的单一联合优化问题。为说明这一事实，为简单起见，我们假设 θ和 h 均为一维参数，且一个二元实值函数 $l(O;\theta,h)$的梯度等于$[m\,(O;\ \theta,h)\quad u\,(O;\ h)]$，或等于 $[u\,(O;\ h)\quad m\,(O;\ \theta,h)]$。于是，利用函数的梯度是保守向量场这一事实，我们得到

$$\frac{\partial m\,(O;\theta,h)}{\partial \theta} = \frac{\partial u(O;h)}{\partial h} \qquad \text{or} \qquad \frac{\partial m\,(O;\theta,h)}{\partial h} = \frac{\partial u(O;h)}{\partial \theta} = 0.$$

例如，若使用第 5.2 小节中的逆概率加权估计量，则可以证明这两个条件均不成立。不过，这两个估计方程可以合并为一个方程组，用以识别参数 θ_0 和 h_0。

5 模拟

5.1 G-估计

我们考虑第 3.1 小节中的模型，并实施算法 4.3。数据生成机制如下：$X \in \mathrm{R}^6$ 服从均值为零、协方差矩阵为 Σ_X 的多元正态分布，其中第 (i, j) 个元素等于 $0.8^{|i-j|}$。随后，我们生成二元处理 Z 和结果 Y，使得

$$Y = \theta_0 Z + g_0(X) + U \tag{7}$$
$$Z|X \sim \text{Bern}(e_0(X)), \tag{8}$$

其中 $\theta_0 = 3, U \sim \mathcal{N}(0, 1)$

$$g_0(X) = X_1 + \exp(X_2 - 1) + |X_3| + \exp(X_4 - 1) + |X_5| + |X_5 X_6|$$
$$e_0(X) = e(X; h_0) = \text{expit}(h_0^\top X)$$

其中 $h_0 = (1, 2, 3, 4, 5, 6)^\top/16$。我们针对不同样本量进行了包含 1000 次重复试验的模拟。在每次迭代中，我们使用矩约束 $\mathrm{E}m(O; \theta_0, h_0) = 0$估计频率学派平均处理效应，其中

$$m(O; \theta, h) = (Y - \theta Z)\{Z - e(X; h)\}.$$

ˆ我们通过含截距的逻辑回归得到 h，并据此估计 h_0。对于每次重复试验，我们实施算法 4.3，并使用 1000 个贝叶斯自助样本推导目标参数的后验分布。我们在表 1 中报告结果，这些结果印证了第 4.3 小节证明的贝叶斯/频率学派二元性。在给出

	n		
	250	500	1000
后验均值的平均值	3.01	2.99	3.00
经验频率学派均值	3.01	2.99	3.00
后验方差的平均值 ($\times n$)	20.89	21.44	21.71
经验频率学派方差 ($\times n$)	23.19	22.27	21.57
三明治估计的平均值	21.99	22.06	21.98
平均贝叶斯可信区间	(2.45, 3.57)	(2.59, 3.39)	(2.71, 3.29)
频率学置信区间	(2。39，3。58)	(2.58, 3.42)	(2。71，3。29)
后验覆盖率	92.6	94.0	94.2

表 1：通过评分 m 和关联贝叶斯自助法估计的平均处理效应的频率学分布与后验分布的比较。平均处理效应的真实值与夹心方差的渐近方差分别为3和22.32

对于可应用“关联贝叶斯自助法”的其他关注的因果量，我们简要说明在贝叶斯自助法程序中使用倾向评分的固定估计、而不考虑其不确定性所产生的影响。结果报告于表2。我们看到，对于 $n = 1000$，平均处理效应的经验后验方差相同，正如理论所表明的那样。就覆盖率而言，仅当h_0 已知时，覆盖率才处于名义水平。在其余情形中，正如定理 4.1之后的分析所表明的，后验分布表现出高于名义水平的覆盖率。

5.2 加权估计量

现在，我们使用由估计方程（6）推导出的逆概率加权估计量，对平均处理效应进行贝叶斯分析。我们生成混杂变量 $X_1 \sim \mathcal{N}(1/2,\ 1/16)$ 和 $X_2 \sim \mathcal{N}(1,\ 1/4)$，并令 $X = (X_1,\ X_2)^\top$。我们根据以下结构模型（7）生成二元处理 Z 和结果 Y，其中 $\theta_0 = 3$、$U \sim \mathcal{N}(0,\ 1)$、$g_0(X) = 2X_1 + X_2$、$e_0(X) = e(X; h_0) = \text{expit}(h_0^\top X)$，且 $h_0^\top = (1,\ 1/2)$。与前一项研究一样，实验针对三种不同样本量各进行 1000 次重复试验。在每次迭代中，我们通过频率学派逻辑回归估计 h_0，并使用逆概率加权估计量估计频率学派的平均处理效应。逆概率加权估计量不依赖结果模型，但其推断有效性完全取决于倾向评分设定正确，且倾向评分不取极限值 0 和 1。

我们在每次迭代中使用1000个贝叶斯自助样本实现联结贝叶斯自助法算法。结果报告于表3，凸显了定理 4.2所证明的贝叶斯/频率学派二元性。为了评估在常规代入法中使用IPW估计的平均处理效应的后验分布的行为，

	n					
	250		500		1000	
h_0	87.33	(94.5)	85.38	(93.2)	83.39	(94.5)
逻辑回归（含截距）	86.55	(99.8)	84.51	(99.9)	82.93	(100.0)
逻辑回归（无截距）	89.57	(96.4)	86.61	(96.0)	83.58	(97.1)
LLB（含截距）	88.46	(99.8)	85.35	(100.0)	83.24	(100.0)
LLB（无截距）	91.32	(96.5)	87.41	(95.9)	83.94	(97.0)

表2：使用固定插入估计量 h_0 时处理效应的经验后验方差平均值（乘以 n），如算法 4.2 所示。括号内为基于 2.5% 和 97.5% 百分位数的后验分布覆盖率。第一行中的标题 h_0 表示使用 h_0 的真实值，而“逻辑回归”表示对 Z 关于 X 进行频率学派逻辑回归。“LLB”表示使用损失似然自助法，其目标函数等于对数似然。后验分布由 500 个贝叶斯自助样本生成，然后取平均。

	n		
	250	500	1000
后验均值的平均值	3.00	3.00	3.00
经验频率学派均值	3.00	3.00	3.00
后验方差的平均值（$\times n$）	6.46	6.01	5.81
经验频率学方差 ($\times n$)	6.04	5.62	5.82
三明治估计的平均值	6.24	5.97	5.82
平均贝叶斯可信区间	(2.68, 3.31)	(2.78, 3.21)	(2.85, 3.15)
频率学置信区间	(2.70, 3.30)	(2.80, 3.20)	(2。86，3。15)
后验覆盖率	95.4	95.3	94.8

表3：分别使用逆概率加权估计量和关联贝叶斯自助法估计平均处理效应，对频率学分布和后验分布进行比较。平均处理效应的真实值以及夹心方差的渐近方差分别为3和5.68。

其中使用固定估计 h_0，我们重复了先前研究中进行的实验。结果报告于表4，与定理 4.1的结论一致。特别是，后验方差不受 h_0 的固定估计影响，且高于在关联贝叶斯自助法下得到的后验方差。此外，正如理论所预测的，这些常规两步法仅在 h_0 已知时才能达到名义水平的覆盖率。加权估计量也可用于估计ATT ρ_0，其满足E$r(O;\rho_0,h_0) = 0$，其中

$$r(O;\rho,h) = ZY - \frac{Y(1-Z)e(X;h)}{1-e(X;h)} - \rho Z.$$

我们进行了模拟研究，对该估计量进行贝叶斯分析。结果列于表5（关联贝叶斯自助法）和表6（传统两步法）。在此模型中，ATT 与平均处理效应相等。

6 讨论

我们推导了通过估计方程定义目标参数所对应的贝叶斯后验分布的性质，并使用贝叶斯自助法进行了计算。为此，我们利用贝叶斯/频率学派二元性的结果，为采用特定计算策略（如关联贝叶斯自助法）提供依据。

在因果场景中，当结果模型的形式未知时，基于倾向评分进行调整是很有价值的，并且正如估计方程理论所示，这有助于一致地估计关注的因果量。当然，我们掌握的治疗分配过程信息仅限于其参数形式。这些假设一直是因果推断中频率学派参数建模的核心，并推动了单重稳健模型的发展；在这类模型中，结果模型的函数形式未知。在贝叶斯框架下，若结果模型设定错误，联合估计倾向评分和关注的因果量可能导致贝叶斯估计出现不理想的行为。结果模型设定错误会反馈至倾向得分模型，这种反馈会妨碍倾向评分发挥平衡作用，并导致处理效应估计有偏。针对反馈问题，人们提出了相应的解决方法，例如截断反馈和两步法，我们已在第4.1小节中简要回顾。所提出的“关联贝叶斯自助法”

	n					
	250		500		1000	
h_0	51.04	(95.5)	51.06	(94.3)	51.44	(93.7)
逻辑回归（含截距）	52.34	(100.0)	51.71	(100.0)	51.75	(100.0)
逻辑回归（无截距）	51.78	(100.0)	51.58	(100.0)	51.67	(100.0)
LLB（含截距）	53.32	(100.0)	52.15	(100.0)	51.95	(100.0)
LLB（无截距）	52.41	(100.0)	51.89	(100.0)	51.81	(100.0)

表4：使用逆概率加权估计量估计平均处理效应时，经验后验方差的平均值（乘以 n），其中 h_0 采用固定插入估计量。括号内为基于第2.5%和第97.5%百分位数计算的后验覆盖率。用于获得 h_0 的估计值 h 的方法与表2中的相同。第一行中的标题 h_0 表示使用了 h_0 的真实值，而“逻辑回归”表示使用 Z 对 X 进行频率学派逻辑回归。“LLB”表示使用损失似然自助法，其目标函数为对数似然。后验分布通过500个贝叶斯自助样本生成，然后取平均。

	n		
	250	500	1000
后验均值的平均值	3.00	3.00	3.00
经验频率学派均值	3.00	3.00	3.00
后验方差的平均值 ($\times n$)	7.55	6.75	6.51
经验频率学派方差 ($\times n$)	7.02	6.36	6.49
三明治估计的平均值	7.19	6.65	6.49
平均贝叶斯可信区间	(2。66，3。32)	(2.76, 3.22)	(2.84, 3.16)
频率学置信区间	(2.68, 3.31)	(2.77, 3.22)	(2。83，3。15)
后验覆盖率	95.4	95.1	95.0

表5：在关联贝叶斯自助法下，使用评分 r 估计的 ATT 的频率学分布与后验分布比较。ATT 的真实值和夹心方差的渐近方差分别为 3 和 6.18。

这是利用贝叶斯自助法近似狄利克雷过程后验分布来理解两步法的另一种视角。“关联”一词强调，倾向评分的不确定性应使用与获取处理效应贝叶斯估计相同的狄利克雷权重进行传播。本文详细研究了关联贝叶斯自助法，并确立了其理想的频率学性质，包括它所满足的贝叶斯/频率学派二元性。这一特定结果是参数型伯恩斯坦–冯·米泽斯定理的一个类比，使我们能够在正确的名义水平上构造贝叶斯可信区间。此外，我们还将关联贝叶斯自助法解释为一种完全贝叶斯的一步法，可对处理和倾向评分进行联合后验推断，且不存在反馈。这一解释在第4.5小节中推导得出，避免了传播倾向得分模型不确定性的问题和插件法这一有问题的概念，并将这两个参数都视为关注参数。最后，我们意识到参数假设带来的建模限制。当然，我们也知道，可以采用高度灵活的方法来估计因果效应，如 Antonelli 等人（2022）、Hahn 等人（2020）和 Vegetabile 等人（2020）所示。这些方法并非本文的重点，而是当前研究工作的主题。

致谢： 作者感谢加拿大自然科学与工程研究委员会（NSERC）和数学科学研究所（ISM）的支持。

补充材料

假设 1（技术假设）。 以下假设（见 Cheng 和 Huang，2010）规定了理论陈述所需的相关概率空间。我们假设

1. 数据是来自概率空间（$\mathcal{O}$，B，P_O）的实现。2. 贝叶斯自助法权重定义在概率空间（$\mathcal{W}$，$\mathcal{C}$，P_W）上。3. 观测值 O_i 是从（$\mathcal{O}^\infty$，B^∞，P_O^∞）到相应坐标空间的典范投影的第 i 个坐标。

	n					
	250		500		1000	
h_0	37.82	(94.7)	37.24	(95.1)	36.31	(95.2)
逻辑回归（含截距）	38.55	(100.0)	37.20	(100.0)	36.54	(100.0)
逻辑回归（无截距）	37.78	(100.0)	37.09	(100.0)	36.50	(100.0)
LLB（含截距）	40.25	(100.0)	37.96	(100.0)	36.90	(100.0)
LLB（无截距）	38.87	(100.0)	37.63	(100.0)	36.75	(100.0)

表6：使用 h_0 的固定插入估计量时，处理组平均处理效应的经验后验方差（乘以 n）的平均值。括号内为基于 2.5% 和 97.5% 百分位数的后验覆盖率。第一行中的标题 h_0 表示使用了 h_0 的真实值，而“逻辑回归”表示使用频率学派逻辑回归对 Z 关于 X 进行回归。“LLB”表示使用损失似然自助法，其目标函数等于对数似然。后验分布通过使用 500 个贝叶斯自助样本生成，然后取其平均值。

4. 来自观测数据和权重的联合随机性定义在乘积概率空间上

$$(\mathcal{O}^\infty, B^\infty, P_O^\infty) \times (\mathcal{W}, \mathcal{C}, P_W) = (\mathcal{O}^\infty \times \mathcal{W}, B^\infty \times \mathcal{C}, P_O^\infty \times P_W).$$

因此，联合概率测度为 $P_{OW} = P_O^\infty \times P_W$。必要时，为简便起见，我们用 P_O 代替 P_O^∞。

假设2 （定理 4.1的假设）**.** 以下假设在实践中并不难以验证。它们通常通过对 m施加条件来验证；参见Tsiatis (2006)。这些假设旨在确保

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(O_i; \tilde{\theta}, \tilde{h})}{\partial \theta^\top} \text{ and } \sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \tilde{\theta}, \tilde{h})}{\partial \theta^\top}$$

依概率收敛于

$$M_{\theta_0} = \mathbb{E}_{P_O}\left\{ \left. \frac{\partial m\,(O; \theta, h_0)}{\partial \theta^\top} \right|_{\theta=\theta_0} \right\},$$

$\tilde{}$ $\tilde{}$前提是 θ 和 h 分别是 θ_0 和 h_0 的相合估计量。关于格利文科–坎特利类和 Donsker 类的可交换自助法详尽论述，可参阅 Praestgaard 和 Wellner（1993）以及 van der Vaart 和 Wellner（1996）。

1. 对某个 $\delta_1,\ \delta_2 > 0$，类 $\left\{ \frac{\partial m}{\partial \theta^\top}(O;\ \theta,\ h) : \|\theta - \theta_0\|_{p,\ 2} < \delta_1,\ \|h - h_0\|_{q,\ 2} < \delta_2 \right\}$ 是 P_O-格利文科–坎特利的，且函数 $(\theta,\ h) \mapsto \mathrm{E}_{P_O}\left\{ \frac{\partial m}{\partial \theta}(O\ \theta\ h) \right\}$；，）连续。

2. 对某个 $\delta > 0$，类 $\left\{ \frac{\partial m}{\partial h^\top}(O;\ \theta_0,\ h) : \|h - h_0\|_{q,\ 2} < \delta \right\}$ 是 P_O-格利文科–坎特利的，且函数 $h \mapsto \mathrm{E}_{P_O}\left\{ \frac{\partial m}{\partial h}(O;\ \ \theta_0\ h) \right\}$，）连续。

3. M_{θ_0} 是一个可逆的 $p \times p$ 矩阵。

定理 *4.1* 的证明。证明的主要思路是建立 $\sqrt{n}(\hat{}\theta_n - \theta_0)$和$\hat{}\ \hat{}\ \sqrt{n}(\theta_{n,\ BB} - \theta_0)$的无条件性质。然后，我们将这两个表达式相减，得到一个关于经验过程 $\sqrt{n}(\mathrm{P}_n^w - \mathrm{P}_n)$的 $\sqrt{n}(\theta_{n,\ BB} - \theta_n)$表达式，以评估其条件和无条件行为。关于以下事实的证明

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{-1}{\sqrt{n}} M_{\theta_0}^{-1} \sum_{i=1}^{n} \{m(O_i; \theta_0, h_0) + M_{h_0}\psi(O_i; h_0)\} + o_{P_O}(1),$$

可以在 Newey 和 McFadden（1994）中找到。为了证明第二个陈述，我们首先证明

$$\sqrt{n}\left(\hat{\theta}_{n,BB} - \theta_0\right) = -M_{\theta_0}^{-1}\left\{ \sqrt{n}\sum_{i=1}^{n} w_{in} m\,(O_i; \theta_0, h_0) + \frac{M_{h_0}}{\sqrt{n}} \sum_{i=1}^{n} \psi(O_i; h_0) \right\} + o_{P_{OW}}(1)$$

我们首先从在 θ_0 附近展开下面的项开始:

$$\sum_{i=1}^{n} w_{in} m\left(O_i; \hat{\theta}_{n,BB}, \hat{h}\right) = \sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, \hat{h}\right) + \left\{\sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \tilde{\theta}, \hat{h})}{\partial \theta^\top}\right\} \left(\hat{\theta}_{n,BB} - \theta_0\right),$$

对于线段上某个 θˆ（该线段连接$\theta_{n,\ BB}$与 θ_0），我们现在来考察

$$\begin{aligned}
&\sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, \hat{h}\right) \\
&= \sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) + \sqrt{n}\left\{\sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \theta_0, \tilde{h})}{\partial h^\top}\right\} \left(\hat{h} - h_0\right) \\
&= \sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) + \left\{\sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \theta_0, \tilde{h})}{\partial h^\top}\right\} \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi\left(O_i; h_0\right) + o_{P_O}(1)\right\}
\end{aligned}$$

对于连接 h 与 h_0的线段上的某个 hˆ，根据假设2，我们有

$$\sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \theta_0, \tilde{h})}{\partial h^\top} = M_{h_0} + o_{P_{OW}}(1) \quad \text{and} \quad \sum_{i=1}^{n} w_{in} \frac{\partial m(O_i; \tilde{\theta}, \hat{h})}{\partial \theta^\top} = M_{\theta_0} + o_{P_{OW}}(1).$$

此外，由于

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(O_i; h_0) = O_{P_{OW}}(1),$$

根据中心极限定理，并结合其不依赖于自助权重这一事实，我们得到

$$\sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, \hat{h}\right) = \sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) + M_{h_0}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(O_i; h_0)\right\} + o_{P_{OW}}(1).$$

此外,

$$\sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) = O_{P_{OW}}(1).$$

综合所有结果，我们得到

$$\begin{aligned}
&\sqrt{n}\left(\hat{\theta}_{n,BB} - \theta_0\right) \\
&= -\left\{M_{\theta_0}^{-1} + o_{P_{OW}}(1)\right\}\left\{o_{P_{OW}}(1) + \sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) + \frac{M_{h_0}}{\sqrt{n}}\sum_{i=1}^{n} \psi(O_i; h_0) + o_{P_{OW}}(1)\right\} \\
&= -M_{\theta_0}^{-1}\left\{\sqrt{n}\sum_{i=1}^{n} w_{in} m\left(O_i; \theta_0, h_0\right) + \frac{M_{h_0}}{\sqrt{n}}\sum_{i=1}^{n} \psi(O_i; h_0)\right\} + o_{P_{OW}}(1).
\end{aligned}$$

令 $\sqrt{n}\left(\hat{\theta}_{n,BB} - \hat{\theta}_n\right) = \sqrt{n}\left(\hat{\theta}_{n,BB} - \theta_0\right) - \sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$，我们得到

$$\sqrt{n}\left(\hat{\theta}_{n,BB} - \hat{\theta}_n\right) = -\sqrt{n} M_{\theta_0}^{-1} \sum_{i=1}^{n}\left(w_{in} - \frac{1}{n}\right) m(O_i; \theta_0, h_0) + o_{P_{OW}}(1).$$

□

假设 3（定理 4.2 的假设）**.** 以下假设与假设2中的假设类似。

1. 类 $\left\{\frac{\partial m}{\partial \theta^\top}(O;\ \theta, h)\right\}$, $\|\theta - \theta_0\|_{p,2} < \delta_1, \|h - h_0\|_{q,2} < \delta_2\Big\}$ 是 P_O-格利文科–坎特利的（对某个 δ_1, $\delta_2 > 0$），且函数 $(\theta, h) \mapsto \mathrm{E}_{P_O}\left\{\frac{\partial m}{\partial \theta}(O\ \theta\ h\ ;\ ,)\right\}$ 是连续的。

2. 该类是 $\left\{\frac{\partial m}{\partial h^\top}(O;\ \theta_0, h),\ \|h-h_0\|_{q,2}<\delta\right\}$ 是 P_O-格利文科–坎特利的（对某个 $\delta>0$），且函数 $h \mapsto \mathrm{E}_{P_O}\left\{\frac{\partial m}{\partial h}(O;\ \theta_0\ h),\right)$ 是连续的。

3. 类 $\left\{\frac{\partial u}{\partial h^\top}(O;\ h),\|h-h_0\|_{q,2}<\delta\right\}$ 对于某个 $\delta>0$ 是 P_O-格利文科–坎特利的，并且函数 $h \mapsto \mathrm{E}_{P_O}\left\{\frac{\partial u}{\partial h}(O\ h\ );\right)$ 是连续的。

4. M_{θ_0} 是一个可逆的 $p\times p$ 矩阵。

5. U_{h_0} 是一个可逆的 $q\times q$ 矩阵。

6. $\hat{h}_n$ 和 $\hat{h}_{n,BB}$ 是 h_0 的无条件相合估计量。

7. $\hat{\theta}_n$ 和 $\hat{\theta}_{n,\ BB}$ 是 θ_0的无条件相合估计量。

定理 *4.2* 的证明。首先，我们注意到

$$\sqrt{n}\left(\hat{h}_n-h_0\right)=-U_{h_0}^{-1}\sqrt{n}\sum_{i=1}^{n}u\left(O_i;h_0\right)+o_{P_O}(1).$$

将上述表达式代入定理 4.1 的第一部分，可得

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)=\frac{-1}{\sqrt{n}}M_{\theta_0}^{-1}\sum_{i=1}^{n}\left\{m(O_i;\theta_0,h_0)-M_{h_0}U_{h_0}^{-1}u(O_i;h_0)\right\}+o_{P_O}(1).$$

现在，我们以与证明 6 类似的方式，研究 $\sqrt{n}\left(\hat{\theta}_{n,\ BB}-\theta_0\right)$。我们注意到

$$\sqrt{n}\left(\hat{h}_{n,BB}-h_0\right)=-U_{h_0}^{-1}\sqrt{n}\sum_{i=1}^{n}w_{in}u\left(O_i;h_0\right)+o_{P_{OW}}(1).$$

通过类似的展开，我们由此得到

$$\begin{aligned}\sum_{i=1}^{n}w_{in}m\left(O_i;\hat{\theta}_{n,BB},\hat{h}_{n,BB}\right)&=\sum_{i=1}^{n}w_{in}m\left(O_i;\theta_0,\hat{h}_{n,BB}\right)\\&\quad+\left\{\sum_{i=1}^{n}w_{in}\frac{\partial m\left(O_i;\tilde{\theta},\hat{h}_{n,BB}\right)}{\partial\theta^\top}\right\}\left(\hat{\theta}_{n,BB}-\theta_0\right)\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}w_{in}m\left(O_i;\theta_0,\hat{h}_{n,BB}\right)&=\sum_{i=1}^{n}w_{in}m\left(O_i;\theta_0,h_0\right)+\left\{\sum_{i=1}^{n}w_{in}\frac{\partial m(O_i;\theta_0,\tilde{h})}{\partial h^\top}\right\}\left(\hat{h}_{n,BB}-h_0\right)\\&=\sum_{i=1}^{n}w_{in}m(O_i;\theta_0,h_0)-M_{h_0}U_{h_0}^{-1}\sum_{i=1}^{n}w_{in}u(O_i;h_0)+o_{P_{OW}}(1)\end{aligned}$$

因此，

$$\sqrt{n}\left(\hat{\theta}_{n,BB}-\theta_0\right)=-M_{\theta_0}^{-1}\sqrt{n}\left\{\sum_{i=1}^{n}w_{in}m(O_i;\theta_0,h_0)-M_{h_0}U_{h_0}^{-1}\sum_{i=1}^{n}w_{in}u(O_i;h_0)\right\}+o_{P_{OW}}(1).$$

通过写成 $\sqrt{n}\left(\hat{\theta}_{n,BB}-\hat{\theta}_n\right)=\sqrt{n}\left(\hat{\theta}_{n,BB}-\theta_0\right)-\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)$，我们得到

$$\sqrt{n}\left(\hat{\theta}_{n,BB}-\hat{\theta}_n\right)=-M_{\theta_0}^{-1}\sqrt{n}\sum_{i=1}^{n}\left(w_{in}-\frac{1}{n}\right)\left\{m\left(O_i;\theta_0,h_0\right)-M_{h_0}U_{h_0}^{-1}u\left(O_i;h_0\right)\right\}+o_{P_{OW}}(1)$$

□

参考文献

Antonelli, J., Papadogeorgou, G. 和 Dominici, F.（2022）。“高维因果推断：贝叶斯建模与良好的频率学派性质的结合。”《生物统计学》，78(1)：100–114。https://doi.org/10.1111/biom.13417。Cheng, G. 和 Huang, J. Z. （ 2010 ）。“ 一般半参数 M 估计的自助法一致性。”《统计年刊》，38(5)：2884 – 2915。Ferguson, T. S.（1973）。“若干非参数问题的贝叶斯分析。”《统计年刊》，1(2)：209 – 230。Ghosal, S. 和 van der Vaart, A. W.（2017）。《非参数贝叶斯推断基础》。剑桥统计与概率数学丛书。剑桥大学出版社。Ghosh, J. K. 和 Ramamoorthi, R. V.（2003）。《贝叶斯非参数方法》。施普林格纽约。Graham, D. J., McCoy, E. J. 和 Stephens, D. A.（2016）。“双重稳健估计的近似贝叶斯推断。” 贝叶斯分析，11(1)：47 – 69。Haeusler, E., Mason, D. M. 和 Newton, M. A.（1991）。“均值的加权自助法。”《*CWI* 季刊》，4(3)：213–228。Hahn, P. R., Murray, J. S. 和 Carvalho, C. M.（2020）。“用于因果推断的贝叶斯回归树模型：正则化、混杂与异质性效应。” 贝叶斯分析，15(3)：965–1056。Hall, P.（1992）。《自助法与埃奇沃思展开》。Springer，纽约。Henmi, M. 和 Eguchi, S.（2004）。“关于干扰参数与投影估计函数的悖论。”《生物统计学》，91(4)：929–941。Hoshino, T.（2008）。“用于潜变量建模的贝叶斯倾向评分调整与马尔可夫链蒙特卡罗算法。”《计算统计与数据分析》，52(3)：1413–1429。Kaplan, D. 和 Chen, J. （ 2012 ）。“ 倾向评分分析的两步贝叶斯方法：模拟与案例研究。”《心理计量学》，77(3)：581–609。Kosorok, M.（2008）。《经验过程与半参数推断导论》。施普林格统计丛书。施普林格纽约。Li, F., Ding, P. 和 Mealli, F.（2023）。“贝叶斯因果推断：批判性综述。”《皇家学会哲学汇刊 *A* 辑：数学、物理与工程科学》，381(2247)：20220153。Lyddon, S. P., Holmes, C. C. 和 Walker, S. G.（2019）。“一般贝叶斯更新与损失似然自助法。”《生物统计学》，106(2)：465–478。Ma, S. 和 Kosorok, M. R.（ 2005）。“稳健半参数 M 估计与加权自助法。”《多元分析杂志》，96(1)：190–217。McCandless, L. C., Douglas, I. J., Evans, S. J. 和 Smeeth, L.（2010）。“在倾向评分贝叶斯回归调整中截断反馈。”《国际生物统计学杂志》，6(2)。McCandless, L. C., Gustafson, P. 和 Austin, P. C.（2009）。“观察性数据的贝叶斯倾向评分分析。”《医学统计学》，28(1)：94–112。Moodie, E. E. M., Saarela, O. 和 Stephens, D. A.（2018）。“处理组平均处理效应的双重稳健加权估计量。”《*Stat*》，7(1)：e205。E205 sta4.205。Moodie, E. E. M. 和 Stephens, D. A.（2022）。“因果推断：过去、现在与未来的关键进展。”《加拿大统计学杂志》，50：1299–1320。Newey, W. K. 和 McFadden, D.（1994）。“大样本估计与假设检验。”载于 Engle, R. F. 和 McFadden, D.（编），《计量经济学手册》，第4卷，第36章，2111–2245。爱思唯尔，第1版。Newton, M. A. 和 Raftery, A. E.（1994）。“使用加权似然自助法进行近似贝叶斯推断。”《皇家统计学会杂志：*B*辑（方法论）》，56(1)：3–48。Praestgaard, J. 和 Wellner, J. A.（1993）。“一般经验过程的可交换加权自助法。”《概率年刊》，21(4)：2053 – 2086。Robins, J. M., Mark, S. D. 和 Newey, W. K.（1992）。“通过对给定混杂变量时暴露期望的建模来估计暴露效应。”《生物统计学》，48(2)：479–495。Rosenbaum, P. R. 和 Rubin, D. B.（1983）。“倾向评分在因果效应观察性研究中的核心作用。”《生物统计学》，70(1)：41–55。

Rubin, D. B. (1981). “贝叶斯自助法。”《统计年刊》，9(1)：130–134。Saarela, O., Stephens, D. A., Moodie, E. E. M., and Klein, M. B. (2015). “关于边际结构模型的贝叶斯估计。”《生物统计学》，71(2)：279–288。Shao, J. and Dongsheng, T. (1995). 刀切法与自助法。Springer，纽约。Stephens, D. A., Nobre, W. S., Moodie, E. E. M., and Schmidt, A. M. (2023). “模型设定错误下的因果推断：基于倾向评分的调整（含讨论）。”贝叶斯分析，18(2)：639 – 694。Tsiatis, A. A. (2006). 半参数理论与缺失数据。施普林格统计丛书。施普林格纽约。van der Vaart, A. W. and Wellner, J. A. (1996). 弱收敛与经验过程：及其在统计学中的应用。施普林格统计丛书。施普林格。Vegetabile, B. G., Gillen, D. L., and Stern, H. S. (2020). “用于估计处理效应的最优平衡高斯过程倾向评分。”《英国皇家统计学会杂志：*A*辑（社会统计学）》，183(1)：355–377。Zigler, C. M. (2016). “贝叶斯定理在因果效应与倾向评分联合估计中的核心作用。”《美国统计学家》，70(1)：47–54。Zigler, C. M., Watts, K., Yeh, R. W., Wang, Y., Coull, B. A., and Dominici, F. (2013). “贝叶斯倾向评分估计中的模型反馈。”《生物统计学》，69(1)：263–273。