arXiv:2603.03569v1 [stat.ME] 3 Mar 2026

通过均值-方差平滑进行细分层下的贝叶斯方差估计

Sepideh Mosaferi *和 Shonosuke Sugasawa †

摘要

细分层调查在许多应用中都很有用，因为其点估计量无偏，但在该设计下，方差估计量难以获得，尤其是每层样本量仅为一个单位时。克服这一困难的一种常见做法是将层两两合并，创建伪层，然后估计方差。由此得到的方差估计量在设计下并非无偏，且当配对伪层的总体均值差异越大时，正偏差也越大。因此得到的置信区间可能会不必要地偏宽。本文提出一种新的贝叶斯方差估计量，不同于文献中已有的方法，它不依赖于合并层。我们采用惩罚样条法，以非参数方式同时平滑均值和方差。此外，我们将该方法与 Breidt 等（2016）的早期研究进行比较。通过多项模拟研究，以及利用全国家庭增长调查（NSFG）数据进行的实例分析，我们展示了该方法的良好表现。

关键词：合并层、基于核的方差估计量、惩罚样条法、初级抽样单位。

*佛罗里达国际大学。通讯作者（电子邮件：smosafer@fiu.edu）
†庆应义塾大学经济学院

重要性声明

本文提出了一种新的贝叶斯方法，用于估计常用精细分层设计的方差，并采用惩罚样条法同时平滑均值和方差。与文献中以往的研究不同，该方法不依赖于合并层。我们提供了两项模拟研究以及一个使用真实数据集的示例，以支持我们的工作并进行比较。

1 引言

在每个互不重叠且可数的子总体（即所谓的层）中，独立抽取极少数单位（通常为一个或两个初级抽样单位，PSU）的过程称为细分层。美国人口普查局的收入与项目参与调查（Westat，2001）和美国农业部的国家资源清查（Nusser 和 Goebel，1997年；Breidt，2002）等调查，分别采用多阶段每层两个单元设计和分层两阶段区域抽样设计。采用三阶段抽样设计的加拿大健康测量调查（Mantel 和 Giroux，2009年）也以每层少量初级抽样单位为基础。由美国人口普查局开展的当前人口调查（人口普查局，2006）和全国犯罪受害调查（Lohr（1999），第 7.6 节），以及密歇根大学社会研究所受美国国家卫生统计中心委托开展的全国家庭增长调查（Lepkowski 等，2010），均采用多阶段每层抽取一个样本的设计。

尽管这种设计广受欢迎，能够使通常用于估计感兴趣参数（例如总体均值或总量）的霍维茨-汤普森（1952）估计量实现无偏且高效，但人们也认识到它存在若干局限。具体而言，每层只有一个初级抽样单位时，设计下的方差不存在；每层有两个或三个初级抽样单位时，

每层的变异则很大。在细分层下估计未知参数方差的一种传统方法，是将相邻的层合并，形成包含更多初级抽样单位的伪层，然后计算方差。该方法最早由 Hansen 等人（1953年）提出，但往往会高估估计量的方差。

本文提出一种贝叶斯方法，作为在细分层情况下估计方差的替代方法。所提出的贝叶斯方法具有以下优点：（i）无需合并层；（ii）同时对各层的均值和方差建模，与现有方法相比，性能显著提升；（iii）易于实施，且除点估计外，还能提供不确定性度量。

1.1 初级抽样单位

在实际应用中，称为初级抽样单位的单位类型取决于调查。初级抽样单位是第一阶段抽样单位。它可能就是最终关注的单位，但通常是某种形式的整群，例如学校、医院或业务单位；在区域概率样本中，它通常由县、人口普查区、街区组或地方行政单位构成。为保持等概率抽选，通常以与其人口规模成比例的概率抽选初级抽样单位，并在第二阶段抽选数量相同的次级抽样单位；但出于其他原因（例如无应答调整），也可以采用不等的纳入概率。例如，在1998年心理健康组织调查（SMHO）中，医院是初级抽样单位。在学校调查中，学校就是初级抽样单位（更多示例见 Valliant 等人（2013）第10.3节）。在本文所述的模拟研究中，我们将抽样单位视为所关注的初级抽样单位。

在区域样本中，层数足够多，因此每层只需抽取一个或两个初级抽样单位。每层抽取一个初级抽样单位，比抽取两个或更多初级抽样单位更能控制最终样本的分布，但会造成一些方差估计问题

本文将对这些问题进行探讨。

1.2 相关研究

文献中提出了几种替代方法，以克服合并层方差估计量的局限性。Hansen 等人（1953年）和 Isaki（1983年）提出使用一些与层均值期望值相关性较强的辅助变量，以降低合并层方差估计量的偏差。Mantel 和 Giroux（2009年）提出了一种基于不同抽样阶段方差分量的方法，并将其应用于加拿大健康测量调查。Mosaferi（2015年）针对每层一个单位的抽样设计的方差，提出了经验贝叶斯估计量和约束经验贝叶斯估计量。作者还比较了每层一个初级抽样单位的设计与每层两个初级抽样单位的设计，并指出，由于采用矩估计法参数估计，所提出的估计量存在一些不稳定性。

最近，Breidt 等（2016）提出了一种非参数方法，用核加权层邻域估计量取代合并层估计量；该方法利用拟合均值函数的偏差来估计方差。他们论文的基本假设是，层均值（或总量）随辅助变量平滑变化。他们提出的方法仍然依赖于合并层。

关于信息性抽样下的贝叶斯方法，Wang 等人（2018年）提出了一种适用于信息性抽样的近似贝叶斯方法，该方法被用于回归估计（Sugasawa 和 Kim，2022年）。此外，Zhao 等人（2020年）在信息性抽样下采用了贝叶斯经验似然。然而，目前尚无专门用于联合估计均值和方差的方法。

1.3 论文结构

本文其余部分的组织如下。第2节介绍一些记号，并回顾基于合并层方法和

基于核的方法的细分层方差。第 3 节针对每个层包含一个初级抽样单位和多个初级抽样单位的情形，提出我们的贝叶斯模型设置，并说明其在实践中的应用方式。第4节通过若干模拟研究展示我们的方法及其性能。此外，在第5节中，我们将该方法应用于来自 NSFG 的一个真实数据集。第6节给出一些结论以及未来研究方向。补充材料包含推导细节和附加模拟结果。

2 精细分层的方差

在分层抽样设计中，我们将目标人群划分为有限个互不相交的层，其中每一层内的单位是同质的，并至少共享一个共同特征，例如位置、收入、教育或性别。这种划分与总体中各层分布已知相结合，使得感兴趣参数的方差相对于简单随机样本有所降低。此外，由于各层相互独立，分层允许对每一层采用灵活的样本选择方法，例如有放回或不放回的简单随机抽样以及系统抽样。在本文中，我们关注精细分层设计这一特殊情形，其中每一层的样本量小至一个初级抽样单位，或为不放回抽取的两个初级抽样单位。我们的目标参数是总体均值，目标是估计估计均值的方差。通常，该方差会带来复杂性，因为当每一层的样本量小至一个单位时，它是不可估计的。在本节其余部分，我们给出一些记号，并解释传统的合并层方差估计量以及基于核的方差方法。

2.1 符号与估计

考虑一个包含 N个单位、指标集为 $U = \{1, 2, ..., N\}$的有限总体，该总体被划分为 H个互不相交的层，例如$U = \cup_{h=1}^{H} U_h$，每层的规模为 N_h，适用于 $h = 1$，2，.....，H

使得 $N = \sum_{h=1}^{H} N_h$。令 y_{hj} 表示嵌套在第 h层中的第 j个个体的取值，使得 $j = 1$，.....，N_h。此外，假设 y_{hj}的取值观测无误。我们从第 h层中随机抽取一个样本，采用不放回抽样，样本量为 n_h，其中 $n_h \geq 1$，且样本总量为 $n = \sum_{h=1}^{H} n_h$。

设 u_h 表示从层 h 中抽取的、样本量为 n_h 的样本，使得总样本为 $u = u_1 \cup u_2 \cup ... \cup u_H$，其中 $u_h \subset U_h$。此外，设 $\pi_j = \mathbb{P}\{j \in u_h\}$ 为一阶入样概率，$\pi_{jk} = \mathbb{P}\{j, k \in u_h\}$ 为二阶入样概率。分层设计下的有限总体均值为 $\bar{y} = N^{-1} \sum_{h=1}^{H} \sum_{j=1}^{N_h} y_{hj}$。分层设计下有限总体均值的霍维茨-汤普森（1952）无偏估计量为

$$\bar{y}_{\text{HT}} = \frac{1}{N} \sum_{h=1}^{H} \sum_{j=1}^{n_h} \frac{y_{hj}}{\pi_j}, \tag{1}$$

其方差为

$$\begin{aligned} V(\bar{y}_{\text{HT}}) &= N^{-2} \sum_{h=1}^{H} S_h^2 \qquad \text{where} \\ S_h^2 &= \sum_{(j,k) \in U_h} \sum \Delta_{jk} (y_{hj}/\pi_j)(y_{hk}/\pi_k) \quad \text{with} \quad \Delta_{jk} = \pi_{jk} - \pi_j \pi_k. \end{aligned} \tag{2}$$

（2）中给出的 $\bar{y}_{\text{HT}}$ 的方差的无偏估计量为

$$var(\bar{y}_{\text{HT}}) = \frac{1}{N^2} \sum_{h=1}^{H} s_h^2 \quad \text{where} \quad s_h^2 = \sum_{(j,k) \in u_h} \sum \frac{\Delta_{jk}}{\pi_{jk}} \frac{y_{hj}}{\pi_j} \frac{y_{hk}}{\pi_k}. \tag{3}$$

在许多应用中，抽样单位 u_h是初级阶段单位。本文仅关注单阶段抽样，并假设采用分层设计（有关该设计的定义，请参见 Valliant 等人（2013）第 3 章）。这种设计很容易推广到更复杂的调查设计，例如多阶段每层两个单元设计，或层内多阶段抽取。

2.2 合并层方差估计量

当每层只有一个初级抽样单位时，无法得到公式（3）中给出的方差估计量，因为 $\pi_{jk} = 0$。此时可以使用合并方差估计量，通常将相邻的两层配对以形成更大的层，从而使每个新的伪层包含更多抽样单位。可以根据空间位置或层规模度量来创建伪层；这些信息通常来自总体框，并且应在样本选择过程开始前确定。为简单起见，假设层的总数“H”为偶数。若 $\{x_h\}$ 表示 $h = 1$、..... H的合并指标，则可根据 x_h对各层排序，从而依次创建伪层。

合并层方差估计量为

$$var_{\text{Coll}} = \frac{1}{N^2}\sum_{h=1}^{H} s^2_{\text{Coll-h}} \quad \text{where} \quad s^2_{\text{Coll-h}} = \frac{1}{2}\Big(\sum_{j\in u_h}\frac{y_{hj}}{\pi_j} - \sum_{l=1}^{H} c_{l(h)}\sum_{k\in u_l}\frac{y_{lk}}{\pi_k}\Big)^2, \tag{4}$$

and

$$c_{l(h)} = \begin{cases} 1 & h \neq l \text{ belong to the same collapsed stratum;} \\ 0 & \text{otherwise.} \end{cases}$$

关于式（4）中给出的合并层方差估计量，另请参见 Breidt 等（2016）、S¨arndal 等（2003）第 109 页，以及 Wolter（2007）第 2.5 节。

方差　　式（4）中的估计量是设计有偏的，其相对于抽样设计的偏差为

$$\text{Bias}(var_{\text{Coll}}) = \frac{1}{2N^2}\sum_{h=1}^{H}\Big(\sum_{j\in U_h} y_{hj} - \sum_{l=1}^{H} c_{l(h)}\sum_{k\in U_l} y_{lk}\Big)^2. \tag{5}$$

偏差项为正；当 $\sum_{j\in U_h} y_{hj} \approx \sum_{k\in U_l} y_{lk}$ for $c_{l(h)} = 1$ 时，它会变小。为进一步简化这一思路，假设 $N_h = N/H$，且每层只抽取一个单位。那么，在不放回简单随机抽样下，偏差项与 $2H^{-2}\sum^{H}{}_{(h\neq l)=1}(\bar{y}_h - \bar{y}_l)^2$ 成正比，其中 $\bar{y}_h = N_h^{-1}\sum_{j\in U_h} y_{hj}$ 和 $\bar{y}_l$ 可用类似方式定义。当 $N \to \infty$ 时，偏差项的阶为 $O(H^{-1})$，这提示我们可以采用一种策略，将各层配成对，以

尽量减小相对于目标特征的偏差幅度，从而最小化差异 $|\bar{y}_h - \bar{y}_l|$。

我们预计，随着层数增加，偏差会减小。当“H”为奇数时，可以将三个原始层（例如）合并为一个伪层，其余伪层则由两个原始层合并而成。随后，我们采用本文所述的相同方法估计方差。Rust 和 Kalton（1987年）比较了通过合并两个、三个或更多层得到的方差估计量。他们指出，合并的层数越多，方差估计量的偏差就越大。因此，成对合并层似乎是一个不错的策略。此外，对于 Breidt 等（2016）提出的基于核的方差估计量，应预先将层配对，尤其是在每层只有一个初级抽样单位的情况下；否则，方差项将无法识别。与文献中的先前研究不同，我们在第 3 节提出的估计量不依赖于合并层。

2.2.1 创建伪层

通常，应在样本选择之前，根据层的特征（如规模、位置、时间结构或层总量（基于抽样框变量））对层进行合并。令 x_h 表示合并指标，该指标对每个层 $h = 1,\ldots, H$均为已知。然后，通过按 x_h 的已知值对层进行排序以形成连续的层对，构造合并函数$c_{l(h)}$ 。层的配对应一次性设定，且不应依赖于所选样本的结果。若 $\bar{y}_h$与 x_h相关，则配对是有效的，且表达式(5)中给出的偏差较小。例如，若使用总体规模或空间位置来形成层，则可以将初级抽样单位规模相近的两个层合并。根据所选样本的特征合并层可能导致方差估计值出现负偏差。

2.3 基于核的方差估计量

Breidt 等（2016）给出的非参数核方差估计量如下

$$var_{\text{Ker}} = \frac{1}{N^2}\sum_{h=1}^{H} s^2_{\text{Ker-h}} \quad \text{where} \quad s^2_{\text{Ker-h}} = \frac{1}{C_d}\Big(\sum_{j\in u_h}\frac{y_{hj}}{\pi_j} - \sum_{l=1}^{H} d_{l(h)}\sum_{k\in u_l}\frac{y_{lk}}{\pi_k}\Big)^2, \tag{6}$$

其中，合并层方差（4）中给出的二元函数 $d_{l(h)}$ 被替换为以下核权重

$$d_{l(h)} = K\Big(\frac{x_h - x_l}{b}\Big)/\sum_{l=1}^{H} K\Big(\frac{x_h - x_l}{b}\Big),$$

使得 $K(.)$ 为 一个对称有界核函数，且 b 为带宽参数。

非随机归一化常数 C_d 为

$$C_d = \frac{1}{H}\sum_{h=1}^{H}\Big(1 - 2d_{h(h)} + \sum_{l=1}^{H} d^2_{l(h)}\Big),$$

它仅取决于核权重。核权重 $d_{l(h)}$ 采用 Epanechnikov 核 $K(u) = 0.75(1-u^2)1_{\{|u|\le 1\}}$。式（6）中给出的核估计量 $s^2_{\text{Ker-h}}$ 存在偏差；参见 Breidt 等（2016）第 3 节。

注释 1. 估计量 (6) 存在偏差，因此，如果各层中的 s^2_h 恒定，应以降低 (3) 中由 s^2_h 引起的偏差为原则来选择 C_d。

注释 2. 对于基于核的方差估计量，带宽应趋于零（$b \to 0$），并满足基于 Breidt 等（2016）中给出的 A6 条件。在其指导原则中，他们建议将带宽选在 $(1/H, 2/H)$ 之间。本文在模拟以及对 NSFG 数据集的示例性分析中，假设 $b = (1/H + 2/H)/2$。

（6）的主要缺点如下：（1）当分布的支撑集并非整个实数轴时，尤其是在

与收入、薪资等相关的应用中，使用对称核会导致权重被分配到观测域之外，从而产生边界偏差（参见 Bouezmarni和Scaillet（2005））；（2）在某些情况下， C_d 可能等于零，导致 var_{Ker} 无法定义；（3）带宽的选择应当谨慎；（4） $d_{l(h)}$ 取决于协变量 $x_h - x_l$，这可能会在某些边界情形下带来困难；（5）该方法仍依赖于合并层。

3 贝叶斯方差估计量

在本节中，我们描述我们的贝叶斯估计量：首先为每一层的均值和方差定义两个未知函数。然后，我们打算在所有层上同时平滑这些函数，并最终估计未知参数以及方差。

3.1 每层一个初级抽样单位的设计

首先，我们考虑一个简单情形，即 $n_h = 1$ （每一层单个单元）。我们对 y_h假设如下模型：

$$y_h \sim N(m(x_h), s^2(x_h)), \quad h = 1, \ldots, H, \tag{7}$$

其中， $m(x_h)$和 $s^2(x_h)$分别是每一层均值和方差的未知函数。我们假设 x_h 是第h层的合并指标或任何其他可用协变量。

为了对未知函数进行建模，我们采用惩罚样条法（例如 Ruppert，2002；Opsomer 等，2008），具体如下：

$$m(x_h;\beta) = \beta_0 + \beta_1 x_h + \cdots + \beta_q x_h^q + \sum_{l=1}^{L} \beta_{q+l}(x_h - \kappa_l)_+^q, \quad \text{and} \tag{8}$$

$$\log s^2(x_h;\gamma) = \gamma_0 + \gamma_1 x_h + \cdots + \gamma_q x_h^q + \sum_{l=1}^{L} \gamma_{q+l}(x_h - \kappa_l)_+^q. \tag{9}$$

在方程 (8) 和 (9) 中，$(\beta_0, \beta_1, \ldots, \beta_q, \beta_{q+1}, ..., \beta_{q+L})$ 和 $(\gamma_0, \gamma_1, \ldots, \gamma_q, \gamma_{q+1}, ..., \gamma_{q+L})$ 是未知系数。此外，$(x)_+^q = \max(x, 0)^q$，而“q”是多项式的次数。这里，我们假设 $\kappa_1, \ldots, \kappa_L$ 为结点， L 是用户指定的结点个数。当结点分布在 x 的整个空间上时，即使 q 较小（例如 $q = 2$ 或 3），样条函数也具有足够的灵活性。尽管我们可以对 $m(x;\beta)$ 和 $s^2(x;\gamma)$ 允许 q 与 L 取不同的值，但为简便起见，我们使用相同的值。特别地，在我们的模拟和 NSFG 示例中，我们假设 $q = 2$ 和 $L = 7$ 。

模型（7）允许我们根据合并指标“x”调整均值和方差参数，因此我们将该方法称为均值-方差平滑。为防止样条函数过拟合，我们假设 $\beta_{q+1}, \ldots, \beta_{q+L}$ 为随机效应，

$$\beta_{q+l} \overset{\text{ind}}{\sim} N(0, \tau_\beta^2), \quad \text{for} \quad l = 1, \ldots, L,$$

其中， τ_β^2 是一个未知方差参数。类似地，我们假设 $\gamma_{q+1}, \ldots, \gamma_{q+L}$ 为随机效应，使得

$$\gamma_{q+l} \overset{\text{ind}}{\sim} N(0, \tau_\gamma^2), \quad \text{for} \quad l = 1, \ldots, L,$$

其中未知参数为 τ_γ^2。

与模型（7）类似，Nott（2006）以及 Pati 和 Dunson（2014）在文献中研究了未知均值函数与方差函数的贝叶斯建模。在此，我们将抽样概率 π_h 的信息纳入后验分布的求取中。为

此，我们考虑以下联合后验分布：

$$\pi(\boldsymbol{\Theta})\prod_{h=1}^{H}\phi(y_h;m(x_h;\beta),s^2(x_h;\gamma))^{\tilde{w}_h}\prod_{l=1}^{L}\phi(\beta_{q+l};0,\tau_\beta^2)\phi(\gamma_{q+l};0,\tau_\gamma^2), \tag{10}$$

其中，$\tilde{w}_h=\pi_h^{-1}/H^{-1}\sum_{h=1}^{H}\pi_h^{-1}$是归一化权重，而 $\phi(\cdot;\ \mu,\ \sigma^2)$是均值为 μ、方差为 σ^2的正态分布的密度函数。

与方程（10）相关的详细步骤见补充材料中的附录 C。虽然后验分布（10）称为一般后验（例如 Bissiri 等，2016），但通过使用满足 $H^{-1}\sum_{h=1}^{H}\tilde{w}_h=1$的归一化权重，可以充分控制其不确定性。$\tilde{w}_h$的存在使结果更加稳定（相关模拟研究见补充材料中的附录 D）。

此外，（10）中的 $\pi(\boldsymbol{\Theta})$对应于未知参数 $\boldsymbol{\Theta}=(\beta_0$、……、$\beta_q$，$\gamma_0$、……、$\gamma_q$，$\tau_\beta^2$，$\tau_\gamma^2)^\top$的先验分布。注意，表达式（10）的对数似然可表示为

$$\sum_{h=1}^{H}\tilde{w}_h\log\phi(y_h;m(x_h;\beta),s^2(x_h;\gamma)),$$

这等价于信息抽样中广为人知的伪似然或加权似然（例如，Godambe and Thompson, 1986；Pfeffermann，1993）。

为了估计 HT估计量 $\bar{y}_{\mathrm{HT}}$ 的方差，我们采用一种混合方法，并将基于设计的权重纳入考量。根据 γ 的后验样本，可以得到 s^2（x_h；γ）的后验样本。因此，可将后验均值用作相应的点估计。为此，HT估计量的贝叶斯方差估计量可表示为

$$var_{\mathrm{Bayes}}=\frac{1}{N^2}\sum_{h=1}^{H}w_h^2\hat{s}^2(x_h;\gamma),$$

其中，$w_h\propto 1/\pi_h$ 是基于抽样设计的相关权重。

Zheng 和 Little（2003年）尝试了另一种方法，他们通过惩罚样条对条件均值 $E(y_j|\pi_j)$ 建模，以解决加权估计量效率低下的问题

并研究了 HT 估计量的替代方案。最终，他们将这些方案与 HT 估计量进行了比较。请注意，我们在方程（7）中对 $y_j|\mathbb{I}\{j \in u_h\}$ 进行了建模，其中 $j = h$ 在此处记作 $n_h = 1$ 。这种建模框架有望适用于任何基于精细分层设计的估计量。

3.1.1 先验设定与后验计算

本节为未知参数向量 $\boldsymbol{\Theta}$ 定义先验分布。对于均值和方差的回归系数，我们分别为 $k = 1, \ldots, p$ 独立指定 $\beta_k \sim N(0, S_\beta)$ 和 $\gamma_k \sim N(0, S_\gamma)$。其中， S_β 和 S_γ 为固定的超参数。此外，对于方差参数，我们指定 τ_β^2 ～逆伽马分布(a_β, b_β) 和 τ_γ^2 ～逆伽马分布(a_γ, b_γ)，其中 a_β, b_β、 a_γ和 b_γ为固定的超参数。这里，逆伽马分布(a, b) 表示密度与 $x^{a-1}\exp(-b/x)$ 成正比的逆伽马分布，其中 $x > 0$。

作为默认选择，我们使用 $S_\beta = S_\gamma = 100$ 和 $a_\beta = b_\beta = a_\gamma = b_\gamma = 1$，从而为回归系数设置弥散先验，并为方差参数设置弱信息先验。式（10）给出的联合后验分布可通过马尔可夫链蒙特卡罗（MCMC）算法进行近似，具体采用 Metropolis-within-Gibbs抽样。全条件分布推导及抽样步骤的详细内容见补充材料。

3.2 每层包含多个初级抽样单位的设计

上一节中的层次模型可以扩展到包含多个初级抽样单位的情形。令 n_h 为第 h层的样本量。接下来，我们考虑以下模型:

$$\boldsymbol{y}_h \sim N_{n_h}(m(x_h)\mathbf{1}_{n_h}, s^2(x_h)\boldsymbol{R}_h(\rho)), \quad h = 1, \ldots, H, \tag{11}$$

其中 $\boldsymbol{y}_h = (y_{h1}, \ldots, y_{hn_h})^\top$，且 $\mathbf{1}_n$是一个 n 维的全 1 向量。此外，表达式 (11) 中的 $\boldsymbol{R}_h(\rho)$ 是一个 $n_h \times n_h$ 矩阵，其 (i, j) 元素为 $\rho^{I(i \neq j)}$；即

$$\boldsymbol{R}_h(\rho) = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix},$$

其中 $\rho \in (-1, 1)$ 是一个未知的相关参数。这里，$m(x_h)$ 和 $s^2(x_h)$ 是关于合并指标 x_h 的标量函数。

对于均值函数 $m(x)$ 和方差函数 $s^2(x)$，我们采用第 3.1 节所讨论的单个 PSU 情形中定义的相同 P样条公式。我们将 $\tilde{w}_{hj}$ 定义如下：

$$\tilde{w}_{hj} = \frac{\pi_{hj}^{-1}}{(\sum_{h=1}^{H}\sum_{j=1}^{n_h}\pi_{hj}^{-1})/\sum_{h=1}^{H} n_h},$$

其中 $\sum_{h=1}^{H}\sum_{j=1}^{n_h}\tilde{w}_{hj}/\sum_{h=1}^{H} n_h = 1$。为了定义后验分布，我们需要定义式 (10) 中使用的幂似然。

设 $\boldsymbol{z}_h = \boldsymbol{R}_h(\rho)^{-1/2}\{\boldsymbol{y}_h - m(x_h, \beta)\mathbf{1}_{n_h}\}$，根据式 (11)，其服从 $N(\mathbf{0}_{n_h}, s(x_h, \gamma)\boldsymbol{I}_{n_h})$。那么，$\boldsymbol{z}_h = (z_{h1}, \dots, z_{hn_h})$ 的幂似然可定义为

$$\prod_{j=1}^{n_h}\phi(z_{hj}; 0, s(x_h,\gamma))^{\tilde{w}_{hj}} \propto s(x_h,\gamma)^{-\sum_{j=1}^{n_h}\tilde{w}_{hj}/n_h}\exp\left\{-\frac{\boldsymbol{z}_h^\top \boldsymbol{W}_h \boldsymbol{z}_h}{2s(x_h,\gamma)}\right\},$$

其中 $\boldsymbol{W}_h = \text{diag}(\tilde{w}_{h1}, \dots, \tilde{w}_{hn_h})$。通过变量替换，$\boldsymbol{y}_h$ 的幂似然与以下表达式成正比

$$\begin{aligned} &L_w(\boldsymbol{y}_h; \beta, \gamma, \rho) \\ &= |\boldsymbol{R}_h(\rho)|^{-1/2} s(x_h,\gamma)^{-\sum_{j=1}^{n_h}\tilde{w}_{hj}/n_h}\exp\left[-\frac{\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)}{2s(x_h,\gamma)}\right], \end{aligned}$$

其中 $\boldsymbol{r}_h(\boldsymbol{y}_h; \beta) = \boldsymbol{y}_h - m(x_h, \beta)\mathbf{1}_{n_h}$。那么，未知

参数的联合后验分布可定义为

$$\pi(\boldsymbol{\Theta}) \prod_{h=1}^{H} L_w(\boldsymbol{y}_h; \beta, \gamma, \rho) \prod_{l=1}^{L} \phi(\beta_{q+l}; 0, \tau_\beta^2) \phi(\gamma_{q+l}; 0, \tau_\gamma^2),$$

其中，$\boldsymbol{\Theta} = (\beta_0、\ldots\ldots、\beta_q、\gamma_0、\ldots\ldots、\gamma_q、\tau_\beta^2、\tau_\gamma^2、\rho)^\top$，且 $\pi(\boldsymbol{\Theta})$ 是其先验分布。

对于除 ρ 之外的参数，我们使用与上一节相同的先验分布。对于 ρ，我们使用均匀先验分布；即 $\rho \sim \text{Uniform}(-1、1)$。我们通过马尔可夫链蒙特卡罗生成后验样本，其中抽样步骤的详细信息见补充材料。最后，根据 γ 的后验样本，可以按照第 3.1 节中的相同方式计算总体方差。

基于 γ 和 ρ 的后验样本，可以得到 $s^2(x_h; \gamma)\boldsymbol{R}_h(\rho)$ 的后验样本。我们使用其后验均值作为相应的点估计。HT估计量的贝叶斯方差估计量可表示为

$$var_{\text{Bayes}} = \frac{1}{N^2} \sum_{h=1}^{H} w_{n_h}^\top \hat{s}^2(x_h; \gamma) \hat{\boldsymbol{R}}_h(\rho) w_{n_h},$$

其中，$w_{n_h}^\top \propto (1/\pi_1、\ldots\ldots、1/\pi_{n_h})$ 是基于抽样设计、适用于多个初级抽样单位的相关权重向量。

4 模拟研究

在本节中，我们使用模拟数据集评估估计量的性能。具体而言，我们生成两个数据集。对于第一个数据集，我们假设超总体服从正态分布；对于第二个数据集，我们假设超总体服从伽马分布。

4.1 高斯超总体

本节的模拟研究中，我们沿用 Breidt 等（2016）的设定来生成有限总体。Salinas 等（2020）此前曾使用这一设定提出一种新的校准估计量。Dahlke 等（2013）、Diana 和 Perri（2012）、Singh（2012）以及 Breidt 和 Opsomer（2008）等一系列论文也使用过这一设定。我们生成一个有限分层总体，其中包含 H 层，每层的规模为 $N_h = NH^{-1} = 60$。对于第 h层，令 $x_h = hH^{-1}$。我们使用以下均值函数

$$m_*(x) = 2\frac{g(x) - \min_{x\in[0,1]} g(x)}{\max_{x\in[0,1]} g(x) - \min_{x\in[0,1]} g(x)},$$

其中 $g(x) = 1 + 2(x - 0.5)$。

为了生成调查变量，我们令 $y_j = m_*(x_h) + e_j$，对于 $j = \{1, ..., N_h\}$，使得 $e_j \sim N(0, \phi^2)$。此外，我们假设 $\phi = \{0.25, 0.5, 5\}$、$H = \{50, 100, 200\}$，并令每层的样本量为 $n_h = 1$，2。我们考虑式（1）中给出的均值 HT估计量，并对以下 3 种不同的方差估计方法进行比较：1）合并方差估计量，记为 var_{Coll}；2）Breidt 等（2016）提出的非参数核方差估计量，记为 var_{Ker}；3）我们提出的贝叶斯方差估计量，记为 var_{Bayes}。

底层抽样设计为不放回简单随机抽样，其中所有单位被抽中的机会均等。为合并各层，我们根据 x_h 的有序取值将相邻层配对。需要注意的是，对于我们提出的贝叶斯估计量，我们没有合并各层。我们开展了一项模拟研究，共进行 $R = 1000$ 次重复模拟，每次重复模拟均包含一次 MCMC 运行，共迭代 10，000 次，预热期为 3000 次。

我们的主要评估标准如下：

- e经验绝对偏差（AB）定义为 $\text{AB}(var_{(.)}) = |\mathbb{E}_R(var_{(.)} - V(\bar{y}_{\text{HT}}))|$，
- 经验均方根误差（RMSE）定义为 RMSE$(var_{(。)}) = \{\mathbb{E}_R(var_{(。)} -$

$V(\bar{y}_{\mathrm{HT}}))^2\}^{1/2}$，并且

- 95% 覆盖概率（CP）定义为 $\mathrm{CP}(var_{(.)}) = \frac{1}{R}\sum_{r=1}^{R} I_r$，其中

$$I_r = \begin{cases} 1 & \text{if} \quad \bar{y} \in \mathrm{CI} \\ 0 & \text{o.w.} \end{cases}$$

以及 $\mathrm{CI} = (\bar{y}_{\mathrm{HT}} - 1.96\sqrt{var_{(.)}}, \quad \bar{y}_{\mathrm{HT}} + 1.96\sqrt{var_{(.)}})$。

$H = 50$ 层的结果见表 1。我们观察到，随着ϕ 的值和初级抽样单位数量增加，贝叶斯方差估计量的绝对偏差和均方根误差均比竞争方法下降得更快，从而能更准确地估计真实方差。此外，基于贝叶斯方差估计量得到的覆盖概率始终更接近名义 95% 水平。附加模拟结果见补充材料。

4.2 HMT总体

为了针对不同于高斯分布的超总体开展模拟研究，我们使用 R 中的 `HMT{PracTools}`，根据伽马分布生成一个总体，其中 $(x_h, \ y_h)$，且 x_h 的取值因单位而异（更多详情见 Valliant等人（2020年））。利用该函数构建各层，使其总量 x大致相同。*HMT*总体 最早由 Hansen 等人（1983年）提出。在此，我们生成 $N = 2000$个观测值，分为 $H = 20$ 层。然后，假设 $x_h = \{x_h - \min(x)\}/\{\max(x) - \min(x)\}$。

我们考察了两种基础抽样设计：不放回简单随机抽样（srswor）和系统抽样；在系统抽样情形下，不同单位的入样概率各不相同。我们采用了第 4.1 节中使用的相同评估标准。结果见表 2 和表 3。与上一节的结果一致，我们发现，与合并方差估计量相比，贝叶斯方差估计量得出的方差估计值更加准确——其绝对偏差和均方根误差更小——尤其是在系统抽样设计下。此外，

表 1：高斯总体中方差估计量的比较。注意，基于核的方法为$b = 0.03$，贝叶斯方法为 $L = 7$。我们有 $H = 50$ 层。

ϕ	初级抽样单位估计量		AB	RMSE	95%覆盖概率
0.25	1	var_{Coll}	3e-04	4e-04	0.931
		var_{Ker}	3e-04	4e-04	0.930
		var_{Bayes}	3e-04	4e-04	0.957
	2	var_{Coll}	7.417e-05	9.304e-05	0.945
		var_{Ker}	1.300e-04	1.684e-04	0.943
		var_{Bayes}	7.159e-05	9.033e-05	0.956
0.5	1	var_{Coll}	0.001	0.001	0.932
		var_{Ker}	0.001	0.001	0.933
		var_{Bayes}	0.001	0.001	0.952
	2	var_{Coll}	3e-04	4e-04	0.944
		var_{Ker}	5e-04	7e-04	0.943
		var_{Bayes}	3e-04	3e-04	0.956
5	1	var_{Coll}	0.111	0.139	0.928
		var_{Ker}	0.110	0.140	0.930
		var_{Bayes}	0.077	0.095	0.930
	2	var_{Coll}	0.030	0.037	0.947
		var_{Ker}	0.055	0.070	0.939
		var_{Bayes}	0.026	0.032	0.945

贝叶斯估计量的覆盖概率在不同情景下更为稳定，且比采用其他方法所得结果更接近名义 95% 水平。

5 使用国家家庭成长调查数据进行的示例分析

全国家庭增长调查（NSFG）是一项关于生育、家庭形成与变迁、计划生育、生殖健康及密切相关主题的调查。它是有关各种生育和家庭主题的全国估计值的重要来源。NSFG 的目标人群包括截至首次

表 2：HMT总体方差估计量的比较。抽样设计为不放回简单随机抽样。请注意，基于核的方法的带宽为 $b = 0.06$ ，贝叶斯方法的带宽为 $L = 7$。

PSU	估计量	AB	均方根误差	95%覆盖概率
1	var_{Coll}	0.044	0.061	0.927
	var_{Ker}	0.041	0.054	0.932
	var_{Bayes}	0.030	0.040	0.912
2	var_{Coll}	0.014	0.019	0.937
	var_{Ker}	0.023	0.032	0.911
	var_{Bayes}	0.011	0.015	0.934

表 3：HMT总体方差估计量的比较。抽样设计为系统抽样。请注意，基于核的方法的带宽为 $b = 0.06$ ，贝叶斯方法的带宽为$L = 7$ 。

PSU	估计量	AB	均方根误差	95%覆盖概率
1	var_{Coll}	0.164	0.175	0.997
	var_{Ker}	0.040	0.051	0.925
	var_{Bayes}	0.035	0.043	0.921
2	var_{Coll}	0.190	0.193	0.999
	var_{Ker}	0.024	0.030	0.905
	var_{Bayes}	0.015	0.024	0.968

调查的联系人，居住在住户中，且其惯常居住地为美国50个州及哥伦比亚特区。

NSFG由国家卫生统计中心（NCHS）设计并实施（更多信息请参见 https://www.cdc.gov/nchs/nsfg/about-nsfg.htm）。2015年-2017年NSFG的抽样设计基于15-49岁住户人口的分层多阶段区域概率样本。该抽样设计相当复杂，感兴趣的读者可查阅相关文档（Documents）。

我们在此使用的公共使用数据集与两年期（2015年–2017年）文件有关。我们使用抽样层（SEST）和整群（SECU）的设计变量，以获得估计值的正确标准误。每层有4个整群，共有18层。我们假设每个整群都是一个初级抽样单位。NSFG 的抽样设计和实地调查由密歇根大学社会研究所实施。女性妊娠文件包含9,553个观测值和248个变量。根据报告的数据，

初级抽样单位总数为 N=2,149。

我们使用了 3 个与研究中受访女性妊娠相关的变量：(1) “AGEPREG”（妊娠年龄）、(2) “EDUCAT”（受教育年限）和 (3) “NBRNALIV”（婴儿数量）。此外，我们还使用了“POVERTY”（贫困线收入）来表示 x，并将其在所有层中进行标准化。为描述我们的方法，我们考虑了两种情景：

- 1初级抽样单位：我们在每层中选取了规模最大的一个初级抽样单位。
- 2 初级抽样单位：我们在每层中选取了规模最大的两个初级抽样单位。

表 4：NSFG 方差估计量比较

PSU	估计量	妊娠年龄	受教育年限	婴儿数量
1	var_{Coll}	1.592e-01	4.499e-02	2.690e-04
	var_{Ker}	1.906e-02	4.835e-03	3.291e-05
	var_{Bayes}	3.412e-04	1.540e-04	7.575e-08
2	var_{Coll}	7.507e-01	2.041e-01	1.250e-03
	var_{Ker}	1.786e-02	2.057e-03	1.754e-05
	var_{Bayes}	1.299e-03	4.242e-04	3.353e-07

接下来，我们假设每个初级抽样单位（PSU）中对应于 y_h 和 x_h 平均值的充分统计量，并根据公式（1）计算这 3 个变量的 HT估计量，以及按照本文所述的 3 种方法计算它们的方差。方差的数值列于表 4 中。

我们观察到，基于所提出的贝叶斯方法得到的方差估计值相对较小。此外，我们使用所有方差计算了 95% 置信区间，结果如图 1 所示。基于所提出方法得到的区间比采用其他方法得到的区间更短。我们强调，该数据集中的调查权重差异较大，且初级抽样单位总数相当多，最终入样概率的范围从 0.04 到接近 0.20，这可能导致后验集中程度提高。我们认为，这种现象是加权似然方法的一般特征

而非所提出方法的根本局限，也是同时最小化均值和方差所带来的结果。

6 讨论

精细分层是一种常用的设计，因为它能够尽可能充分地进行分层。本文针对精细分层设计，提出了一种基于惩罚样条法的贝叶斯方差估计量。该方法能够对所有层的均值和方差同时进行平滑。与文中介绍的其他方法不同，我们的方法不依赖合并层，政府机构也可以轻松使用。该建模框架具有灵活性，可扩展至每层存在多个辅助变量的情形，这将留待未来研究。

在实际应用中，初级抽样单位（PSU）通常是由若干单位组成的群体，例如医院或学校。因此，y_h 可作为每个初级抽样单位的充分统计量，与 NSFG 示例类似；我们可以假设，如方程（7）和（11）所述，这些统计量的集合服从正态性假设。轻微至中度的偏离（轻度偏度或尾部略重于正态）通常对标准推断程序影响甚微，尤其是随着层数因中心极限定理（CLT）而增加时。然而，当偏离严重时，例如极端偏度、多峰性或重尾，该假设并不稳健，特别是在层数较少时。对于这些情况，可以改变正态性假设，但我们的主要方法仍保持不变。

虽然使用 $\boldsymbol{R}_h(\rho)$ 来纳入多个初级抽样单位之间的相关性可能很有用，但第 3.2 节所考虑的等相关性假设主要是出于数学上的便利。即使相关结构设定有误，该假设仍可能带来一些好处，这与广义估计方程中的工作相关类似（Liang和Zeger，1986）。如果我们掌握了有关相关结构的一些先验知识，就可以

使用其他相关矩阵，例如自回归相关性。

尽管本文提出的贝叶斯框架主要用于获得方差参数的点估计，但完整的后验分布包含可在未来工作中加以利用的额外信息。具体而言，方差估计值中的后验不确定性可用于刻画方差估计量本身的变异性，这与调查抽样和方差估计中常用的有效自由度概念密切相关。在经典近似下，方差估计量通常与卡方分布相联系，其相应的自由度决定了置信区间和假设检验的准确性。所提方法的一个有趣扩展是，基于后验变异性推导近似的自由度度量，并将其与现有的基于设计的方法或合并层方法进行比较。此类进展可进一步提升所提方法在下游推断中的实用性，包括置信区间的构建和正式统计检验。

最后，所提出的方法聚焦于单变量方差估计，其中对对数方差函数的建模能够以直接的方式保证非负性。未来工作的一个重要方向是将该框架扩展到多变量情形，例如对多个HT估计量进行联合方差估计。在此类情形中，对协方差分量直接进行基于样条的建模，即使所有边际方差均为正，也可能无法自动保证所得协方差矩阵的正定性。解决这一问题需要额外的结构约束。可行方法包括通过结构化参数化对协方差矩阵进行建模，例如自回归先验或随机游走先验，或采用在构造上强制正定性的变换。另一种选择是，基于核的方法或协方差估计的加权平均方法在使用非负权重组合正定的层水平贡献时能够保持正定性。在贝叶斯框架内探索并比较这些策略，是未来研究中一条有前景的途径。

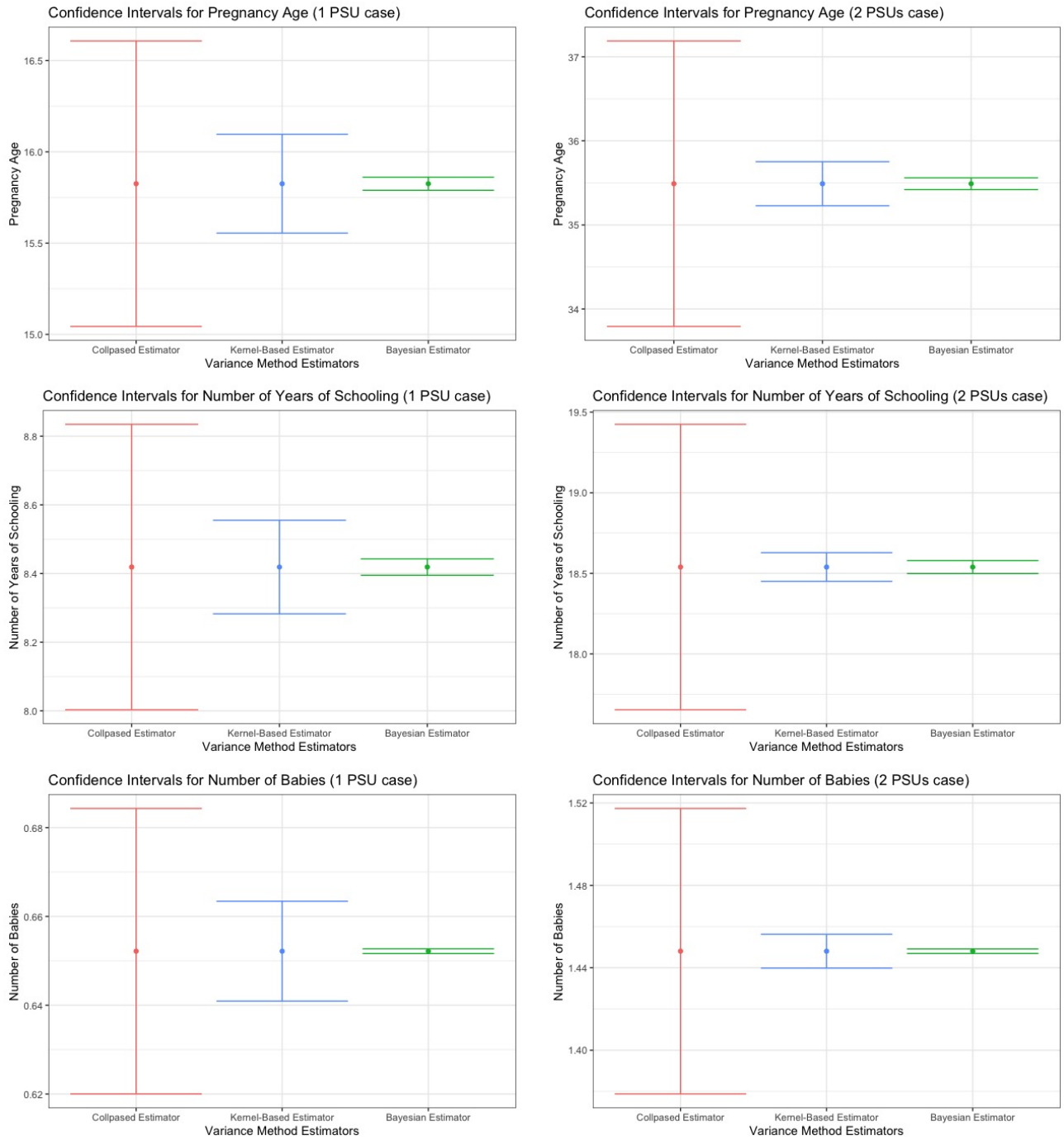

图 1：NSFG 中多个变量的置信区间。请注意，为了提高图表的可读性，HT估计量按 7 倍缩放。

参考文献

Bissiri, P. G., Holmes, C. C. 和 Walker, S. G. (2016) “更新信念分布的一般框架，” 英国皇家统计学会会刊*B*辑：统计方法学，78, 1103–1130。

Bouezmarni, T. 和 Scaillet, O. (2005) “非对称核密度估计量和平滑直方图的一致性及其在收入数据中的应用，” 计量经济学理论，21, 390–412。

Breidt, F. J. (2002) “国家资源清查，” 环境计量学百科全书，135–136。

Breidt, F. J. 和 Opsomer, J. D. (2008) “调查中的内生后分层：使用样本拟合模型进行分类，” 统计年鉴，36, 403–427。

Breidt, F. J., Opsomer, J. D. 和 Sanchez-Borrego, I. (2016) “细分层下的非参数方差估计：合并层的替代方法，” 美国统计学会会刊，111, 822–833。

人口普查局 (2006) 当前人口调查设计与方法技术报告第 66 号，华盛顿特区。

Dahlke, M., Breidt, F. J., Opsomer, J. D. 和 Van Keilegom, I. (2013) “非参数内生后分层估计”，《中华统计学志》23，189–211。

Diana, G. 和 Perri, P. F. (2012) “敏感数据的基于校准的方法：一项模拟研究”，应用统计学杂志39，53–65。

Godambe, V. 和 Thompson, M. E.（1986）“超总体与调查总体的参数：它们的关系与估计”，国际统计评论，127– 138。

Hansen, M. H.、Hurwitz, W. N. 和Madow, W.G.（1953年）抽样调查方法与理论 第一卷 John Wiley。

Hansen, M. H.、Madow, W. G. 和 Tepping, B. J.（1983年）“对样本调查中模型依赖推断与概率抽样推断的评价”，美国统计学会会刊，78，776–793。

Horvitz, D. G. 和 Thompson, D. J.（1952年）“有限总体不放回抽样的一种推广”，美国统计学会会刊， 47，663–685。

Isaki, C. T.（1983年）“利用辅助信息进行方差估计”，美国统计学会会刊，78，117–123。

Lepkowski, J. M.、Mosher, W. D.、Davis, K. E.、Groves, R. M. 和 Van Hoewyk, J.（2010年）“2006—2010年全国家庭成长调查：抽样设计与连续调查分析”，生命与健康统计。第*2*辑，数据评估与方法研究，1–36。

Liang, K.-Y. 和 Zeger, S. L.（1986）“使用广义线性模型进行纵向数据分析”，*Biometrika*，73，13–22。

Lohr, S. L.（1999）《抽样：设计与分析》美国加利福尼亚州太平洋丛林市：Duxbury Press。

Mantel, H. 和 Giroux, S.（2009年）“每层一个初级抽样单位的复杂抽样调查中的方差估计”，华盛顿特区：*JSM*会议论文集，统计计算分部。

Mosaferi, S.（2015年）“每层一个单位的抽样设计下的经验贝叶斯与约束经验贝叶斯方差估计”，美国统计学会调查研究方法分部会议论文集，2452–2464。

Nott, D.（2006）“非高斯数据的均值函数与方差函数半参数估计”，*Computational Statistics*，21，603–620。

Nusser, S. M. 和 Goebel, J. J.（1997年）“国家资源清查：一项长期多资源监测计划”，环境与生态统计，4，181–204。

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G. 和 Breidt, F. J. (2008) “使用惩罚样条回归进行非参数小区域估计，”《皇家统计学会会刊*B*辑：统计方法学》，70,265–286。

Pati, D. 和 Dunson, D. B. (2014) “具有变动残差密度的贝叶斯非参数回归，”统计数理研究所年刊，66, 1–31。

Pfeffermann, D. (1993) “对调查数据建模时抽样权重的作用，”国际统计评论，317–337。

Ruppert, D. (2002) “为惩罚样条选择结点数量，”计算与图形统计学杂志，11, 735–757。

Rust, K. 和 Kalton, G. (1987) “合并层以进行方差估计的策略，”《官方统计学杂志》，3, 69–81。

Salinas, V. I., Sedory, S. A. 和 Singh, S. (2020年) “使用幂变换进行校准，”统计学通讯：模拟与计算，49, 2256–2286。

S¨arndal, C.E., Swensson, B. 和 Wretman, J. (2003年) 模型辅助抽样调查 施普林格科学与商业媒体。

Savitsky, T. D. 和 Toth, D. (2016) “信息性抽样下的贝叶斯估计，”电子统计学杂志，10,1677–1708。

Singh, S. (2012) “使用位移函数校准设计权重，”*Metrika*，75, 85–107。

Sugasawa, S. 和 Kim, J. K.（2022年）“使用许多辅助变量进行模型辅助调查估计的近似贝叶斯方法”，《中华统计学志》，32，477–498。

Valliant, R.、Dever, J. A.和 Kreuter, F.（2013）《调查样本设计与加权实用工具》施普林格。

Valliant, R.、Dever, J. A. 和 Kreuter, F.（2020年）“practools”软件包。

Wang, Z.、Kim, J. 和 Yang, S.（2018年）“信息性抽样下的近似贝叶斯推断”，*Biometrika*，105，91–102。

Westat（2001年）收入与项目参与调查用户指南（第3版），技术报告，马里兰州罗克维尔。

Williams, M. R. 和 Savitsky, T. D.（2021年）“复杂抽样下伪贝叶斯推断的不确定性估计”，国际统计评论，89，72–107。

Wolter, K.（2007年）方差估计导论 施普林格科学与商业媒体。

Zhao, P.、Ghosh, M.、Rao, J. 和 Wu, C.（2020年）“复杂调查数据的贝叶斯经验似然推断”，《皇家统计学会会刊*B*辑：统计方法学》，82，155–174。

Zheng, H. 和 Little, R. J.（2003年）“基于惩罚样条模型的估计：从与规模成比例概率抽样样本中估计有限总体总量”，《官方统计学杂志》19，99–117。

补充材料：

“精细分层下基于均值—方差平滑的贝叶斯方差估计娱乐代理

分层下通过均值-方差平滑进行贝叶斯方差估计”

本补充材料的结构如下。A 节提供了与条件分布相关的一些详细推导。B 节提供附加模拟结果。C 节介绍伪似然的细节，D 节则通过一些模拟研究说明其对结果的影响。

A 详情 的马尔可夫链蒙特卡罗全条件分布

A.1 每层一个初级抽样单位

注意，联合后验分布表示为

$$\pi(\boldsymbol{\Theta})\prod_{h=1}^{H}\phi(y_h;m(x_h;\beta),s^2(x_h;\gamma))^{\tilde{w}_h}\prod_{l=1}^{L}\phi(\beta_{q+l};0,\tau_\beta^2)\phi(\gamma_{q+l};0,\tau_\gamma^2).$$

详细的抽样步骤如下：

- （从 τ_β^2 抽样） τ_β^2 的全条件分布为逆伽马分布$(a_\beta+L/2,b_\beta+\sum_{l=1}^{L}\beta_{q+l}^2/2)$。- （从 τ_γ^2 抽样） τ_γ^2 的全条件分布为逆伽马分布 $(a_\gamma+L/2,b_\gamma+\sum_{l=1}^{L}\gamma_{q+l}^2/2)$。- （从 β 抽样） β 的全条件分布为 $N(A_\beta^{-1}B_\beta,A_\beta^{-1})$，其中

$$A_\beta=\mathrm{diag}(S_\beta^{-1}I_{q+1},\tau_\beta^{-2}I_L)+\sum_{h=1}^{H}\frac{\tilde{w}_h z_h z_h^\top}{s^2(x_h;\gamma)},\qquad B_\beta=\sum_{h=1}^{H}\frac{\tilde{w}_h z_h y_h}{s^2(x_h;\gamma)}$$

with $z_h = (1, x_h, \ldots, x_h^q, (x-\kappa_1)_+^q, \ldots, (x-\kappa_L)_+^q)$.

- （从 γ 抽样） γ 的全条件分布的对数密度（不含无关常数）为

$$\log p_f(\gamma) \equiv \sum_{h=1}^{H} \tilde{w}_h \log \phi(y_h; m(x_h;\beta), s^2(x_h;\gamma)) - \frac{1}{2}\sum_{k=0}^{q}\frac{\gamma_k^2}{S_\gamma} - \frac{1}{2}\sum_{l=1}^{L}\frac{\gamma_{q+l}^2}{\tau_\gamma^2},$$

为此，我们采用随机游走 Metropolis-Hastings 算法。给定当前值，记为 $\gamma^\dagger$，我们从固定方差为 b 的 $N(\gamma^\dagger, bI_{q+L})$ 生成提议值 γ^*，并以概率 $\min\left\{1, p_f(\gamma^*)/p_f(\gamma^\dagger)\right\}$ 接受该提议值。

A.2 每个层中的多个 PSU

注意，联合后验分布表示为

$$\pi(\boldsymbol{\Theta}) \prod_{h=1}^{H} L_w(\boldsymbol{y}_h; \beta, \gamma, \rho) \prod_{l=1}^{L} \phi(\beta_{q+l}; 0, \tau_\beta^2)\phi(\gamma_{q+l}; 0, \tau_\gamma^2),$$

其中

$$L_w(\boldsymbol{y}_h; \beta, \gamma, \rho) = |\boldsymbol{R}_h(\rho)|^{-1/2} s(x_h,\gamma)^{-\sum_{j=1}^{n_h}\tilde{w}_{hj}/n_h} \exp\left[-\frac{\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)}{2s(x_h,\gamma)}\right],$$

且 $\boldsymbol{r}_h(\boldsymbol{y}_h;\beta) = \boldsymbol{y}_h - m(x_h,\beta)\mathbf{1}_{n_h}$。详细的抽样步骤如下：

- （从 τ_β^2 和 τ_γ^2 抽样） τ_β^2 和 τ_γ^2 的全条件分布与单个 PSU 情形下的完全相同。

- （抽样 g 来自 β） β 的全条件分布为 $N(A_\beta^{-1}B_\beta, A_\beta^{-1})$，其中 e

$$A_\beta = \mathrm{diag}(S_\beta^{-1}I_{q+1}, \tau_\beta^{-2}I_L) + \sum_{h=1}^{H}\frac{\boldsymbol{z}_h\mathbf{1}_{n_h}^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\mathbf{1}_{n_h}\boldsymbol{z}_h^\top}{s^2(x_h;\gamma)},$$

$$B_\beta = \sum_{h=1}^{H}\frac{\boldsymbol{z}_h\mathbf{1}_{n_h}^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\mathbf{1}_{n_h}\boldsymbol{y}_h}{s^2(x_h;\gamma)}.$$

- （从 γ 抽样） γ 的全条件分布的对数密度（不含无关常数）为

$$\log p_f(\gamma) \equiv -\frac{1}{2}\sum_{h=1}^{H}\frac{\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)}{2s(x_h,\gamma)} - \frac{1}{2}\sum_{h=1}^{H}\sum_{j=1}^{n_h}\tilde{w}_{hj}\log s(x_h,\gamma) - \frac{1}{2}\sum_{k=0}^{q}\frac{\gamma_k^2}{S_\gamma} - \frac{1}{2}\sum_{l=1}^{L}\frac{\gamma_{q+l}^2}{\tau_\gamma^2},$$

对此，我们采用随机游走 Metropolis-Hastings 算法。给定当前值，例如 $\gamma^\dagger$，我们从 $N(\gamma^\dagger, b_\gamma I_{q+L})$ 中生成一个提议值 γ^*，其方差固定为 b_γ，并以概率 $\min\left\{1, p_f(\gamma^*)/p_f(\gamma^\dagger)\right\}$ 接受该提议值。

- （从 ρ 抽样） ρ 的全条件分布的对数密度（不含无关常数）为

$$\log p_f(\rho) \equiv -\frac{1}{2}\sum_{h=1}^{H}\frac{\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)^\top \boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{W}_h\boldsymbol{R}_h(\rho)^{-1/2}\boldsymbol{r}_h(\boldsymbol{y}_h;\beta)}{2s(x_h,\gamma)} - \frac{1}{2}\sum_{h=1}^{H}\log|\boldsymbol{R}_h(\rho)|,$$

对此，我们再次使用随机游走 Metropolis-Hastings 算法。给定当前值（记为 $\rho^\dagger$），我们从 $N(\gamma^\dagger,\ b_\rho)$ 中生成一个提议值 ρ^*，其中方差固定为 b_ρ，并以概率 $\min\left\{1, p_f(\rho^*)/p_f(\rho^\dagger)\right\}$ 接受该提议值。

B 附加模拟结果

本节给出高斯超总体的附加模拟结果。在表 B.1 和 B.2 中，我们分别给出 $H = 100$ 层和 $H = 200$ 层的模拟结果。结果表明， var_{Bayes} 估计量的表现可能显著优于替代估计量。此外，在表 B.3 中，我们考察了方差估计量 var_{Bayes} 在 $L =$ 的不同取值（5，7，10）下的表现。我们观察到， L 取不同值时，结果并无显著差异。

表 B.1：高斯总体方差估计量之间的比较。注意，对于基于核的方法， $b = 0.015$ ；对于贝叶斯方法， $L = 7$ 。我们有$H = 100$ 层。

ϕ	PSU	估计量	AB	RMSE	95%覆盖概率
0.25	1	var_{Coll}	9.386e-05	1e-04	0.942
		var_{Ker}	9.550e-05	1e-04	0.948
		var_{Bayes}	9.076e-05	1e-04	0.953
	2	var_{Coll}	2.752e-05	3.402e-05	0.955
		var_{Ker}	4.593e-05	5.797e-05	0.958
		var_{Bayes}	2.558e-05	3.194e-05	0.950
0.5	1	var_{Coll}	4e-04	5e-04	0.943
		var_{Ker}	4e-04	5e-04	0.948
		var_{Bayes}	3e-04	4e-04	0.950
	2	var_{Coll}	1.100e-04	1e-04	0.955
		var_{Ker}	1.837e-04	2e-04	0.958
		var_{Bayes}	1e-04	1e-04	0.950
5	1	var_{Coll}	0.038	0.047	0.942
		var_{Ker}	0.038	0.049	0.947
		var_{Bayes}	0.028	0.035	0.937
	2	var_{Coll}	0.011	0.014	0.955
		var_{Ker}	0.018	0.023	0.958
		var_{Bayes}	0.010	0.012	0.944

表B.2：高斯总体中各方差估计量的比较。注意，对于基于核的方法，$b = 0.0075$；对于贝叶斯方法，$L = 7$。我们有$H = 200$层。此处未给出每层包含2个初级抽样单位的结果，因为计算量过大。

ϕ	PSU	估计量	AB	RMSE	95%覆盖概率
0.25	1	var_{Coll}	3.348e-05	4.180e-05	0.950
		var_{Ker}	3.337e-05	4.187e-05	0.947
		var_{Bayes}	2.708e-05	3.471e-05	0.948
0.5	1	var_{Coll}	1e-04	2e-04	0.950
		var_{Ker}	1e-04	2e-04	0.947
		var_{Bayes}	1e-04	1e-04	0.947
5	1	var_{Coll}	0.013	0.017	0.950
		var_{Ker}	0.013	0.017	0.947
		var_{Bayes}	0.009	0.012	0.943

表 B.3：贝叶斯方法在 $L =$ 取不同值时的性能（5、7、10）。

ϕ	H	PSU	AB	RMSE	95%覆盖概率
0.25	50	1	(3e-04,3e-04,3e-04)	(4e-04,4e-04,4e-04)	(0.961,0.957,0.970)
		2	(7.249e-05,7.159e-05,6.833e-05)	(9.184e-05,9.033e-05,8.572e-05)	(0.954,0.956,0.955)
	100	1	(8.888e-05,9.076e-05,8.865e-05)	(1e-04,1e-04,1e-04)	(0.956,0.953,0.966)
		2	(2.481e-05,2.558e-05,2.462e-05)	(3.109e-05,3.194e-05,3.122e-05)	(0.958,0.950,0.943)
	200	1	(2.879e-05,2.708e-05,2.689e-05)	(3.601e-05,3.471e-05,3.410e-05)	(0.954,0.948,0.961)
0.5	50	1	(0.001,0.001,0.001)	(0.001,0.001,0.001)	(0.955,0.952,0.964)
		2	(3e-04,3e-04,3e-04)	(3e-04,3e-04,3e-04)	(0.951,0.956,0.955)
	100	1	(3e-04,3e-04,3e-04)	(4e-04,4e-04,4e-04)	(0.954,0.950,0.958)
		2	(9.703e-05,1e-04,9.168e-05)	(1e-04,1e-04,1e-04)	(0.958,0.950,0.943)
	200	1	(1e-04,1e-04,1e-04)	(1e-04,1e-04,1e-04)	(0.953,0.947,0.959)
5	50	1	(0.073,0.077,0.072)	(0.091,0.095,0.091)	（0.943,0.930,0.943）
		2	(0.025,0.026,0.024)	(0.032,0.032,0.030)	(0.942,0.945,0.947)
	100	1	(0.027,0.028,0.027)	(0.034,0.035,0.035)	(0.947,0.937,0.948)
		2	(0.009,0.010,0.009)	(0.012,0.012,0.012)	(0.954,0.944,0.942)
	200	1	(0.010,0.009,0.010)	(0.012,0.012,0.012)	(0.948,0.943,0.953)

C 伪似然的详细说明

假设从概念上说， $\phi(\boldsymbol{y};\mu,\ \sigma^2)$ 是总体完整似然。那么，后验密度可写为

$$\pi(\boldsymbol{\Theta},\boldsymbol{\Psi}|\boldsymbol{y},\boldsymbol{x}) \propto \pi(\boldsymbol{\Theta})\,\phi(\boldsymbol{y};\mu,\sigma^2)\prod_{l=1}^{L}\phi(\beta_{q+l};0,\tau_\beta^2)\phi(\gamma_{q+l};0.\tau_\gamma^2),$$

其中， $\boldsymbol{\Psi}=(\beta_{q+1},\ \ldots\ldots,\ \beta_{q+L},\ \gamma_{q+1},\ \ldots\ldots,\ \gamma_{q+L})$。对总体分层并在每层抽取 1 个初级抽样单位后，根据 $h=1,\ \ldots\ldots,\ H$ 的观测样本（$y_h,\ x_h$），我们对数据进行了建模。观测样本的似然可表示为

$$\prod_{h=1}^{H}\phi(y_h;m(x_h;\beta);s^2(x_h;\gamma)).$$

使用后验密度的代入估计量（Savitsky 和 Toth，2016），可得如下结果：

$$\hat{\pi}(\boldsymbol{\Theta},\boldsymbol{\Psi}|\boldsymbol{y},\boldsymbol{x}) \propto \pi(\boldsymbol{\Theta})\prod_{h=1}^{H}\phi(y_h;m(x_h;\beta);s^2(x_h;\gamma))\prod_{l=1}^{L}\phi(\beta_{q+l};0,\tau_\beta^2)\phi(\gamma_{q+l};0.\tau_\gamma^2).$$

然后，为了体现每个样本观测的影响，我们纳入抽样权重，基于样本加权似然来近似总体似然，如下所示

$$\hat{\pi}(\boldsymbol{\Theta},\boldsymbol{\Psi}|\boldsymbol{y},\boldsymbol{x}) \propto \pi(\boldsymbol{\Theta})\prod_{h=1}^{H}\phi(y_h;m(x_h;\beta);s^2(x_h;\gamma))^{\tilde{w}_h}\prod_{l=1}^{L}\phi(\beta_{q+l};0,\tau_\beta^2)\phi(\gamma_{q+l};0.\tau_\gamma^2),$$

其中 $\tilde{w}_h$ 是归一化权重。此外，考虑 $\tilde{w}_h$ 可以让结果比忽略它们时更加稳定。同样，如果每层选择的初级抽样单位多于 1 个，我们可以采用相同的策略处理“去相关”的观测值，$\boldsymbol{z}_h=\boldsymbol{R}_h(\rho)^{-1/2}\{\boldsymbol{y}_h-m(x_h,\ \beta)\mathbf{1}_{n_h}\}$，其权重为 $\tilde{w}_{hj}$，如第 3.2 节所述。

D 伪似然对结果的影响

本节基于小规模模拟研究展示两个示例，探讨在似然中忽略权重的影响。我们遵循稿件中描述的模拟设计，并考虑两种情景：（a）采用不放回简单随机抽样设计并选择 1 个初级抽样单位的高斯情形；（b）采用系统抽样设计并选择 1 个初级抽样单位的 HMT 情形。

我们仅生成 100 个抽样值，并比较 4 种方差估计量；对于最后一种方差估计量“var_{Bayes} (ignorance)”，我们忽略似然中的权重。结果汇总于图 D.1 和表 D.4。根据这些结果，忽略似然中的权重会使结果变得相当不稳定（尤其是在采用系统抽样设计的 HMT 情形下）。因此，将似然中的权重纳入考虑可能会带来一些益处，这与文献中 Savitsky和Toth（2016）以及 Williams和Savitsky（2021）所得的结论一致。

表 D. 4：基于 100 次模拟抽样的方差估计量比较 s.

设计	判据	var_{Coll}	var_{Ker}	var_{Bayes}	var_{Bayes} (无知)
高斯分布	AB	0.090	0.103	0.073	0.073
（包含 1 个初级抽样单位的不放回简单随机抽样）	RMSE	0.117	0.135	0.088	0.089
HMT	AB	0.161	0.046	0.037	0.072
（包含 1 个初级抽样单位的系统抽样）	RMSE	0.175	0.057	0.045	0.168

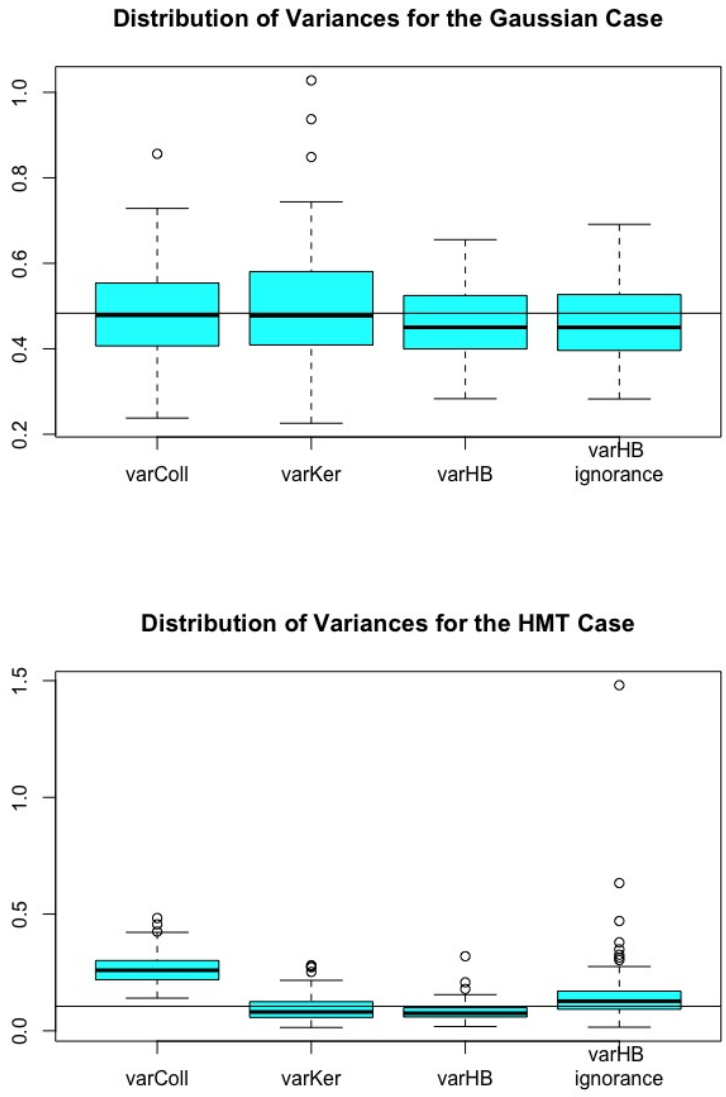

图 D.1：基于 100 次模拟抽样的方差估计量分布比较。水平线表示 HT 估计量 $V(\bar{y}_{\mathrm{HT}})$ 的真实方差。