贝叶斯层次模型的群组变异参数先验设置探讨

喻　瑞[1]　杨绍坤[1]　何从源[1]　李佳蔚[1]　Tianxin Shi[2]　赵　星[1]　郭　冰[1△]

【提　要】 **目的**　本研究旨在探讨不同先验分布对贝叶斯层次模型关键参数估计的影响,并分析群组水平数量变化对估计精度的作用。**方法**　通过案例分析和模拟研究对模型中群组变异参数分别赋予逆伽马分布、半柯西分布和指数分布,比较其性能及受群组水平数量变化的影响。**结果**　指数分布由于尾部较短,可能低估方差;而逆伽马和半柯西分布因为尾部较厚,能提供更为稳健的估计。在群组水平较多的情况下,模型对变异的估计更加精确,而随着群组数量的减少,估计误差会增大。**结论**　在小样本条件下,半柯西和指数分布因其便于结合外部信息且能够合理反映变异参数的变化特征,成为更适当的选择。这些发现为贝叶斯层次模型在实际应用中选择群组变异参数的先验提供了重要参考依据。

【关键词】 贝叶斯层次模型　先验模型　群组变异

【中图分类号】 R195.1　　【文献标识码】 A　　**DOI**　10.11783/j.issn.1002-3674.2024.05.004

Prior Distributions for Variance Parameters in Bayesian Hierarchical Models

Yu Rui, Yang Shaokun, He Congyuan, et al (*Department of Epidemiology and Biostatistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu* 610041)

【**Abstract**】 **Objective**　This study aims to explore the impact of different prior distributions on the estimation of key parameters in Bayesian hierarchical models and to analyze how changes in the number of group levels affect estimation accuracy. **Method**　Through case analysis and simulation studies, we assigned inverse gamma, half-Cauchy, and exponential distributions to the group variance parameters in the model, comparing their performance and the influence of the number of group levels. **Results**　The exponential distribution, due to its shorter tail, may underestimate variance, while the inverse gamma and half-Cauchy distributions, with their thicker tails, provide more robust estimates. In scenarios with a higher number of group levels, the model estimates variance more accurately, but as the number of groups decreases, estimation errors increase. **Conclusion**　In small sample conditions, the half-Cauchy and exponential distributions are more suitable choices because they easily incorporate external information and reasonably reflect the variation characteristics of parameters. These findings provide important guidance for selecting priors for group variance parameters in practical applications of Bayesian hierarchical models.

【**Key words**】 Bayesian hierarchical models; Prior distributions; Group-level variation

层次模型(hierarchical model)是一种处理具有多层结构数据的统计模型。层次模型通过其多层次架构,能够捕捉不同层级之间的变异性和协同效应,借力识别不同亚组或群体中的效应差异[1]。无论是在医疗领域预测未测试药物在特定人群中的效果[2],还是在市场研究中预测未调查地区的消费者行为,层次模型都展示了其强大的适用性和预测能力[3]。

层次模型的估计通常在贝叶斯框架下实现,将先验信息与当前数据相结合,形成后验分布。这种结合不仅增强了模型的解释力,还能在数据稀疏或不完整的情况下提供更稳健的估计。根据先验信息是否用到专业背景知识或其他信息,一般被分为三类:信息先验、弱信息先验和无信息先验[4]。其中,弱信息先验用于在缺乏强烈先验知识的情况下提供适度的约束,以避免过度影响后验分布。对于变异参数,常用的先验包括逆伽马分布、半柯西分布和指数分布[5]。不同的先验分布可能会对后验分布产生不同的影响,探索不同先验对结果的影响可以提高对模型结果的解释力,帮助识别稳健的结果,理解模型在不同假设下的表现。

层次模型中具有群组变异这个特殊参数。群组变异参数估计过程中,与组内样本往往较多不同,群组数目可能不多,使得群组变异参数的有效估计面临挑战。本研究通过应用不同先验分布于真实数据的贝叶斯层次模型,探讨不同先验分布对群组变异参数估计结果的影响。采用统计模拟研究,详细分析了贝叶斯层次模型在不同群组样本量中的适用性和表现差异。通过对比和分析不同先验分布的效果,本研究旨在为贝叶斯层次模型的群组变异参数的先验分布设定提供指导,为实际应用中选择最优的先验分布参数设置提供一定依据。

原理与方法

1.贝叶斯层次结构模型

当存在多个群组,不同群组有多个个体时,贝叶斯层次结构模型可以用如下形式表示:

$$y_i \sim N(\alpha_{j[i]}, \sigma_y^2), i=1,\cdots,n \qquad (1)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), j=1,\cdots,J \qquad (2)$$

1.四川大学华西公共卫生学院 四川大学华西第四医院流行病与卫生统计学系(610041)

2.University of California, Berkeley, College of Engineering

△通信作者:郭冰,E-mail:guobing0111@ scu.edu.cn

其中,j 表示第 j 个群组,i 表示第 i 个个体,$j[i]$ 表示属于第 j 个群组的第 i 个个体;α_j 表示第 j 个群组的平均水平,σ_y 表示同一群组内不同个体的组内变异,μ_α 表示整体平均水平,σ_α 表示不同群组的组间变异。

2.先验设置

现对贝叶斯层次模型中未知参数 $\mu_\alpha,\sigma_\alpha,\sigma_y$ 设置先验。其中,组间变异参数 σ_α 为正数,需要取值为正的概率分布进行描述,文献常用的分布如表 1 所示。

表 1 组间变异参数常用先验分布及其特点

常用分布	特点
逆伽马分布	1. 逆伽马分布是正态分布方差的共轭先验,先验参数具有先验样本的解释性。 2. 劣势在于无标准差的直接解释,其常见参数设置无法描述层次模型中群组变异。
半柯西分布	1. 半柯西分布是最长尾的 t 分布,对于超大的标准差有更强的先验概率强度。 2. 尺度参数就是其中位数,便于通过选择合适值获得弱信息先验。
指数分布	1. 指数分布的短尾特性限制了异常大的取值,适用于参数较为集中、极大取值可能性小的场景。 2. 唯一参数的倒数既是均值也是标准差,便于通过选择合适值获得弱信息先验。

实例研究

1.数据来源

本研究案例数据来自公开数据集 radon data,可通过 R 包 SDAResouces[6] 或 fcbCI[7] 获取该数据。该数据为美国环境保护署所测量的明尼苏达州 85 个县(该州共有 89 个县)的 919 间房屋的氡浓度,具有“县-房屋”的二级嵌套结构。氡是一种致癌物质,它是天然存在的放射性气体,其衰变产物也具有放射性,已知高浓度的氡会导致肺癌,据估计在美国每年造成数千例肺癌死亡。

2. 模型与先验

案例数据中具备典型的嵌套结构:85 个县作为第一层嵌套,不同县内房屋的氡含量检测作为第二层嵌套。我们将对模型中的 $\mu_\alpha,\sigma_\alpha,\sigma_y$ 三个参数设置不同的先验,探讨不同先验设置对参数估计结果的影响(表 2)。

对于 μ_α,设定其服从正态分布,$\mu_\alpha \sim N(\mu_0,\tau_0^2)$,其中 μ_0 和 τ_0 的取值根据文献报道的全球氡含量确定,本文中分别设定为 log(1.3) 和 log(5)[8]。此外,还构造无信息先验(设定 $\mu_0=0,\tau_0=1000$)作为对比。

针对 σ_α,σ_y 我们将设定最为常用的三类先验分布,分别为逆伽马分布、半柯西分布和指数分布。类似地,先验分布中的参数将根据文献报道设定,并且也将进一步设定弱信息先验和无信息先验。通过探索先验的设置规律,我们将共计设立 8 个模型和对应的 6 个敏感性分析模型。

表 2 $\mu_\alpha,\sigma_\alpha,\sigma_y$ 三个参数先验模型设置

模型	μ_α 超参数设定	σ_α、σ_y 超参数设定		参数设定解释
M1	$\mu_\alpha \sim N(\mu_0,\tau_0^2)$	逆伽马分布	$\sigma_\alpha,\sigma_y \sim IG(\frac{v_0}{2},\frac{v_0\sigma_0^2}{2})$	$\mu_0,\tau_0,v_0,\sigma_0$ 根据文献中数据进行设定,整体弱信息先验
M2	$\mu_\alpha \sim N(\mu_0,\tau_0^2)$	逆伽马分布	$\sigma_\alpha,\sigma_y \sim IG(0.001,0.001)$	μ_0,τ_0 根据文献中数据进行设定,σ_α,σ_y 设定为 Jeffery 先验的近似
M3	$\mu_\alpha \sim N(0,1000^2)$	逆伽马分布	$\sigma_\alpha,\sigma_y \sim IG(\frac{v_0}{2},\frac{v_0\sigma_0^2}{2})$	v_0,σ_0 均根据文献中数据进行设定,μ_α 设定为一个较宽泛的无信息先验
M4	$\mu_\alpha \sim N(0,1000^2)$	逆伽马分布	$\sigma_\alpha,\sigma_y \sim IG(0.001,0.001)$	μ_α 设定为一个较宽泛的无信息先验,σ_α,σ_y 设定为 Jeffery 先验的近似
M5	$\mu_\alpha \sim N(\mu_0,\tau_0^2)$	半柯西分布	$\sigma_\alpha,\sigma_y \sim half\text{-}cauchy(\tau_\alpha)$	$\mu_0,\tau_0,\tau_\alpha,\tau_y$ 均根据文献中数据进行设定,整体弱信息先验
M6	$\mu_\alpha \sim N(0,1000^2)$	半柯西分布	$\sigma_\alpha,\sigma_y \sim half\text{-}cauchy(\tau_\alpha)$	μ_α 设定为一个较宽泛的无信息先验
M7	$\mu_\alpha \sim N(\mu_0,\tau_0^2)$	指数分布	$\sigma_\alpha,\sigma_y \sim exp(\lambda_\alpha)$	$\mu_0,\tau_0,\lambda_\alpha,\lambda_y$ 均根据文献中数据进行设定,整体弱信息先验
M8	$\mu_\alpha \sim N(0,1000^2)$	指数分布	$\sigma_\alpha,\sigma_y \sim exp(\lambda_\alpha)$	μ_α 设定为一个较宽泛的无信息先验

:另设定几个敏感性分析的模型,模型 1 中 $v_0=0.001$,模型 3* 中 $v_0=0.001$,整体为极弱信息先验。模型 5* 中 $\tau_\alpha=\tau_y=1000$,模型 6* 中 $\tau_\alpha=\tau_y=1000$;模型 7* 中 $\lambda_\alpha=\lambda_y=0.0001$;模型 8* 中 $\lambda_\alpha=\lambda_y=0.0001$,模型中的 σ_α,σ_y 均无信息先验。

3. 结果展示

案例分析结果显示见表 3,当县个数为 85、μ_α 采用弱信息和无信息正态分布时,σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果一致,采用近似 Jeffery 先验的结果相对偏小但一致。当县个数为 30、μ_α 采用弱信息和无信息正态分布时,σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果差别不大,采用近似 Jeffery 先验的结果相对偏小但一致。当县个数为 10 和 3、μ_α 采用弱信息和无信息正态分布时,σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果差别不大,采用近似 Jeffery 先验的结果偏小且不一致,尤其是当 μ_α 也采用无信息先验时。

表 3 不同先验不同群组样本量下参数 σ_α 的后验分布中位数

模型	$N=85$	$N=30$	$N=10$	$N=3$
M1	0.314	0.342	0.336	0.677
M2	0.307	0.325	0.228	0.207
M3	0.314	0.347	0.348	0.651
M4	0.306	0.325	0.081	0.074
M5	0.311	0.337	0.344	0.744
M6	0.313	0.346	0.336	0.653
M7	0.313	0.339	0.322	0.699
M8	0.313	0.337	0.343	0.684

模拟研究

1. 模拟场景与评价指标

在模拟研究中，本研究进一步探索县水平数和先验分布设置对群组水平均值 μ_α、群组水平变异参数 σ_α、个体水平变异参数 σ_y 三个参数的影响。具体而言，在 1 次模拟中，先从实例研究中的 85 个县随机抽取 30、10、3 个县生成 3 个模拟数据集；其次，分别拟合表 2 中的 8 个模型和 6 个敏感性分析模型；然后，采用马尔科夫蒙特卡洛模拟（1000 次预热和 5000 次的模拟采样）算法对参数进行估计。最后，重复 100 次模拟，计算三个参数的 MSE，即：

$$\mathrm{MSE} = \sum_{i=1}^{n} (\widehat{\beta_i} - \beta)^2 / n$$

其中，n 表示 100 次模拟，β_i 表示第 i 次参数估计值，β 表示 85 个县 14 个模型估计得到的参数的平均值。通过 MSE 来评价模型参数估计的稳健性。

2. 模拟结果

模拟结果见表 4。对于参数 σ_α，当县个数为 30 和 10，μ_α 采用弱信息和无信息正态分布时，σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果基本一致且准确度较高，采用 Jeffery 先验的结果基本一致但准确度较低；当县个数为 3，μ_α 采用弱信息和无信息正态分布时，σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果差别不大但准确度较高，采用近似 Jeffery 先验的结果差别不大但准确度较低。

对于参数 μ_α，当县个数为 30 和 10，σ_α、σ_y 采用相同先验分布时，μ_α 采用弱信息正态分布与采用无信息正态分布的结果基本一致；当县个数为 3，σ_α、σ_y 采用相同先验分布时，μ_α 采用弱信息正态分布比采用无信息正态分布的结果偏小。

对于参数 σ_y，当县个数为 30、10 和 3 时，μ_α、σ_α、σ_y 分别采用不同先验设置的结果差别不大，但随着县个数的减少，准确度逐渐下降。

模拟研究敏感分析结果显示，对于参数 σ_α，当县个数为 30、10 和 3 且 μ_α 采用相同先验时，σ_α、σ_y 采用极弱信息逆伽马分布、半柯西分布和指数分布分别比 σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果大，且随着县个数的减少，所有模型估计的准确度逐渐下降；而对于参数 μ_α 和 σ_y，当县个数为 30、10 和 3 时，σ_α、σ_y 采用极弱信息逆伽马分布、半柯西分布和指数分布分别与 σ_α、σ_y 采用弱信息逆伽马分布、半柯西分布和指数分布的结果基本一致，但随着县个数的减少，准确度逐渐下降。

表 4　不同先验和群组样本量下不同参数的 MSE

模型	μ_α			σ_α			σ_y		
	$N=30$	$N=10$	$N=3$	$N=30$	$N=10$	$N=3$	$N=30$	$N=10$	$N=3$
M1	0.0100	0.0173	0.0563	0.0163	0.0219	0.0331	0.0061	0.0129	0.0406
M2	0.0098	0.0235	0.0756	0.0426	0.0975	0.1428	0.0057	0.0125	0.0418
M3	0.0104	0.0190	0.0723	0.0168	0.0221	0.0333	0.0061	0.0129	0.0407
M4	0.0097	0.0215	0.0748	0.0410	0.0942	0.1501	0.0056	0.0126	0.0397
M5	0.0100	0.0174	0.0548	0.0170	0.0226	0.0386	0.0061	0.0130	0.0427
M6	0.0104	0.0190	0.0728	0.0169	0.0231	0.0446	0.0061	0.0130	0.0428
M7	0.0098	0.0172	0.0544	0.0177	0.0245	0.0412	0.0060	0.0129	0.0426
M8	0.0103	0.0190	0.0729	0.0180	0.0240	0.0456	0.0060	0.0130	0.0430
M1*	0.0090	0.0243	0.0680	0.0414	0.0971	0.1487	0.0056	0.0124	0.0390
M3*	0.0101	0.0228	0.0811	0.0415	0.0999	0.1551	0.0057	0.0124	0.0420
M5*	0.0091	0.0190	0.0588	0.0419	0.0831	0.1178	0.0057	0.0125	0.0403
M6*	0.0095	0.0203	0.0645	0.0435	0.0839	0.1169	0.0057	0.0125	0.0407
M7*	0.0099	0.0173	0.0479	0.0163	0.0214	0.2255	0.0060	0.0128	0.0444
M8*	0.0104	0.0191	0.0767	0.0159	0.0213	0.4274	0.0060	0.0129	0.0443

*：N 表示县数目

讨　论

本研究探讨了不同先验分布的设置对贝叶斯层次模型中群组水平均值 μ_α、群组水平变异参数 σ_α、个体水平变异参数 σ_y 三个参数估计的影响，并探讨了群组水平数量的变化对估计精度的影响。在 BUGS 软件的早期帮助文档中，由于逆伽马分布的共轭特征与近似 Jeffery 先验需求，两个参数都设定为非常接近 0 的逆伽马分布被推荐用于方差的先验[9]，该设置在当前文献依旧较为常见。但其在群组水平上表现可能不佳。从结合外部信息，合理构造先验分布规律角度讲，半柯西分布与指数分布都更为适合。

群组水平的变异参数 σ_α 估计的准确度随县水平数的减少而降低,群组较少时信息先验显得较为重要。当群组样本量(即本例中的县数)较多时,σ_α 能够得到较为准确的估计;随着县水平数的减少,均方误差会升高。贝叶斯层次模型虽然可以很好地利用不同数据层次水平的信息,但是当群组样本数过于稀少时,对于群组水平变异的估计不可避免地有较大的误差,相关研究也有类似的结论[10-11]。在这种情况下,先验信息可以通过提供额外的约束和方向,帮助模型进行更准确的估计。先验信息越丰富,模型对群组方差的估计越稳健,特别是在面对数据稀疏或群组样本量较小的情形时,这一优势更加明显。因此,在对群组水平的变异进行建模时,基于公开文献、经验数据或历史研究结果来设定合适的先验信息显得至关重要[12-13]。通过这些来源,我们可以获取关于方差范围、可能的群组差异等方面的合理期望值,从而提高模型对实际情况的匹配度。这一点在群组样本量较少时尤为关键,因为数据无法单独支撑有效的推断,而恰当的先验信息能够极大地补充这一不足,保证估计结果的合理性和准确性。

当先验分布设置为指数分布时,相较于逆伽马分布和半柯西分布,σ_α 估计的准确度较差。原因可能是指数分布的重尾特性较弱,导致其倾向于产生较小的变异估计值[14-15]。这意味着,当实际的变异较大时,指数分布可能会施加过大的收缩,最终导致 σ_α 低估。这种收缩效应在数据较为稀疏或组间变异较大时显得尤为明显。此外,指数分布对极端值的包容性较差,这限制了它在处理变异较大或异常波动数据时的灵活性。相比之下,半柯西分布具有更厚的尾部,能够更好地捕捉和容忍极端值,从而提供更为稳健的估计[14, 16]。因此,模拟研究表明,指数分布在复杂场景下表现欠佳,逆伽马分布则因其更灵活的特性,能够提供更可靠的估计。

个体变异 σ_y 和整体均值 μ_α 的估计准确对于群组水平的先验设置不敏感,结果稳健。群组水平变异 σ_α 的先验无论是那种情况,在组内数据较为充足时,σ_y 和 μ_α 的估计误差仍然在可接受的范围内。在实际建模过程中,个体层面的数据量往往远大于群组层面的数据,这使得我们能够对个体层面的变异性和群组均值进行较为准确的推断,即使群组的样本量开始缩减,误差的增幅也较为有限。相反,群组水平变异 σ_α 受到样本量的影响更为敏感[17],尤其是在群组样本量较小的情况下,估计容易受到较大的不确定性影响。这也进一步强调了在群组样本量较小时,设定合适的先验信息对于 σ_α 估计的重要性。

综上所述,本研究探讨了贝叶斯层次模型先验分布设置对参数 μ_α、σ_α 和 σ_y 的估计影响,特别关注群组样本量对估计精度的影响。研究表明,群组水平变异 σ_α 的估计在群组样本量较少时依赖于合适的先验分布设定,半柯西分布表现出更好的稳健性。先验信息在群组样本量不足时尤为重要,有助于提高模型的稳健性和估计精度。同时,个体水平的变异 σ_y 和群组均值 μ_α 在样本量充足时通常能得到较为准确的估计,尽管群组样本量减少会引入误差,但整体仍在可接受范围内。这进一步凸显了在群组数据稀疏时,合理选择先验分布对于确保模型准确性的关键作用。

参 考 文 献

[1] Lawler PR, Goligher EC, Berger JS, et al. Therapeutic Anticoagulation with Heparin in Noncritically Ill Patients with Covid-19[J]. The New England Journal of Medicine, 2021, 385: 790-802.

[2] Newman JD, Anthopolos R, Mancini GBJ, et al. Outcomes of Participants With Diabetes in the ISCHEMIA Trials[J]. Circulation, 2021, 144: 1380-1395.

[3] 于石成, 廖加强, 于姝, 等. 复杂抽样数据多水平模型分析方法及其应用[J]. 中国卫生统计, 2014, 31: 193-196,201.

[4] Gelman A, Carlin JB, Stern HS, et al. Bayesian Data Analysis [M]. London: CHAPMAN & HALL/CRC, 2013.

[5] Röver C, Bender R, Dias S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis[J]. Res Synth Methods, 2021, 12: 448-474.

[6] Lu YL, Sharon. R Companion for Sampling: Design and Analysis [M]. Boca Raton, FL: CRC Press, 2021.

[7] Yu C, Hoff PD. Adaptive multigroup confidence intervals with constant coverage[J]. Biometrika, 2018, 105: 319-335.

[8] Chambers DB, Zielinski JM. Residential Radon Levels Around the World[C]//N. JEROME. Encyclopedia of Environmental Health (Second Edition). Oxford: Elsevier, 2011: 508-519.

[9] Spiegelhalter DT A, Best N, Lunn D. WinBUGS User Manual: Version 1.4[M]. Cambridge, UK MRC Biostatistics Unit 2003.

[10] Gelman A. Prior distributions for variance parameters in hierarchical models(Comment on an Article by Browne and Draper)[J]. Bayesian Analysis, 2006, 1: 515-533.

[11] Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models[J]. Bayesian Analysis, 2006, 1.

[12] Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics[J]. British Journal of Mathematical and Statistical Psychology, 2013, 66: 8-38.

[13] McElreath R. Statistical Rethinking: A Bayesian Course with Examples in R and STAN[M]. New York: Chapman and Hall/CRC, 2020.

[14] Andrew Gelman JH. Data Analysis Using Regression and Multilevel/Hierarchical Models [M]. Cambridge UK: Cambridge University Press, 2007.

[15] Kruschke JK. Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan[M]. Amsterdam: Academic Press, 2015.

[16] David Lunn CJ, Nicky Best, Andrew Thomas, et al. The BUGS Book[M]. London: CRC Press, 2012.

[17] Vasishth S, Yadav H, Schad DJ, et al. Sample Size Determination for Bayesian Hierarchical Models Commonly Used in Psycholinguistics[J]. Computational Brain & Behavior, 2022, 6: 102-126.

(责任编辑:郭海强)