

Bootstrap模型平均置信区间

郭庆光, 李新民

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2021年10月11日; 录用日期: 2021年11月1日; 发布日期: 2021年11月16日

摘要

模型平均作为能够解决模型不确定性的统计推断方法越来越受欢迎, 模型平均应用也从点估计扩展到区间估计。目前模型平均置信区间常用方法包括Wald区间和MATA-Wald区间, 这些区间都是在参数估计的分布渐近正态的情况下构建的, 所以这两种方法在拟合正态数据时效果是非常好的。在本文中, 我们将参数估计的Bootstrap分布来近似其真实分布, 进而提出了Bootstrap模型平均置信区间, 并对这3种方法分别在正态数据和偏态数据的模型下的表现进行模拟比较。模拟结果表明在正态数据的模型下, Bootstrap模型平均置信区间在上错误率和覆盖率方面仅略差于MATA区间。在偏态数据的模型下, 百分位数Bootstrap模型平均置信区间的覆盖率要优于Wald区间和MATA区间, 并且提供了较好的错误率, 特别是上错率, 都更接近名义水平。所以Bootstrap区间表现出很好的适应性。

关键词

模型平均, Bootstrap, 上、下错误率, 覆盖率, 平均区间长度

Model-Averaged Bootstrap Confidence Interval

Qingguang Guo, Xinmin Li

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Oct. 11th, 2021; accepted: Nov. 1st, 2021; published: Nov. 16th, 2021

Abstract

Model averaging is becoming more and more popular as a statistical inference method to solve model uncertainty, and the application of model averaging is also extended from point estimation to interval estimation. At present, the common methods of model averaged confidence interval include Wald interval and MATA-Wald interval. These intervals are constructed under the condition

that the distribution of parameter estimation is asymptotically normal. Therefore, these two methods are very good in fitting normal data. In this paper, the bootstrap distribution of parameter estimation was approximated to the real distribution, and then the model-averaged bootstrap interval was proposed. The performance of the three methods under normal and skewed data models was simulated and compared respectively. The simulation results show that the model-averaged bootstrap interval is only slightly worse than that of the MATA model in the upper error rate and coverage under the normal data model, Under the model with skewed data, the coverage of the model-averaged bootstrap interval is better than that of the Wald interval and the MATA interval, and it also provides a better error rate, especially the upper error rate, which is closer to the nominal level. So the bootstrap interval shows good adaptability.

Keywords

Model Averaging, Bootstrap, Upper and Lower Error Rate, Coverage Rate, Mean Interval Length

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

传统上, 参数估计是基于单一模型的假设, 该模型的选择是一个模型选择的过程。这种方法忽视了模型的不确定性[1], 导致参数估计的过拟合。

在许多应用领域进行模型参数估计时, 考虑模型不确定性的情况越来越普遍。这种不确定性之所以会出现, 是因为我们不能确保模型选择过程总是会选择相同的最佳模型。传统上, 经过模型选择过程挑选出最佳模型之后, 再使用挑选出的最佳模型来进行参数估计。该参数的估计依赖于这个最佳模型的选择。由于模型选择的方法一般依赖于挑选出的最佳模型, 这样就会丢弃其他模型所特有的信息。这种方法忽视了模型的不确定性。为了解决模型选择带来的这些问题, 需要考虑在某种程度上接受所有的候选模型。考虑到这个问题的重要性, 人们开始深入学习并研究模型平均方法, 其作为一种简单、通用、易理解的方法, 应用于多个模型的研究过程。这解决了出现的模型不确定性问题。模型平均方法的出现起初更多应用在点估计上, 后来随着理论研究的深入, 区间估计也开始使用这种方法。

模型平均有 Beyes 模型平均[2]和频率模型平均[3]两方面, 本文主要介绍频率模型平均。在频率方法的设置中, 模型平均估计 $\hat{\theta}$ 定义为单个模型估计的加权和: $\hat{\theta} = \sum_{i=1}^R \omega_i \hat{\theta}_i$ 其中 $\hat{\theta}_i$ 是模型 M_i 下的估计, 模型权重 ω_i 是由信息准则决定的, 例如 AIC [4], 和式是基于 R 个候选模型的集合。

目前已经提出了一些构造频率模型平均置信区间的方法。形式为 $\hat{\theta} \pm z_{\alpha} \cdot \widehat{se}(\hat{\theta})$ 的 Wald 区间, 其中 z_{α} 是标准正态分布的 $1-\alpha$ 分位数。其依赖于 $\hat{\theta}$ 的标准差 $se(\hat{\theta})$ 的精确估计, Hjort 和 Claeskens [5]表明 $se(\hat{\theta})$ 估计的困难性, 以及 Wald 区间在覆盖率方面表现很差。Buckland [6] [7]等人提出了 $\widehat{se}(\hat{\theta})$ 的一些选择。

另一个可选的构造频率模型平均置信区间的方法是 MATA-Wald [8]方法, 由 Turek 和 Fletcher [7]提出。在 MATA 方法中, 每个置信限度定义为满足单个模型 Wald 区间错误率的和等于期望错误率的结果, MATA 在小样本时的覆盖率方面要强于现有的模型平均的方法。Fletcher 和 Turek [9]将 MATA 的构造方式应用到轮廓似然区间中, 提出了模型平均轮廓似然区间, 轮廓似然区间在偏态数据的拟合上表现出明显的优势。Kabaila [10] [11] [12]等人研究了 MATA 置信区间的一些性质。Yu [13]等人提出了 MATA 区

间的转换版本。

本文利用单模型下参数的 Bootstrap 置信区间以及类比 MATA-Wald 模型平均置信区间的构造原理, 构造了 Bootstrap 模型平均置信区间, 通过模拟计算来说明其优势。第二部分, 介绍了 Wald 区间的构造方法以及相应的标准差的估计、MATA-Wald 模型平均置信区间的方法及 Bootstrap 模型平均置信区间的构造方法。第三部分, 分别在响应变量是正态和偏态的情形下进行了模拟研究, 从上下错误率、覆盖率和平均区间长度分别比较了 Wald 区间、MATA 区间和 Bootstrap 区间的表现性能。第四部分, 对本文的研究进行了总结。

2. 方法及思想

假设考虑 R 个候选模型 $\{M_1, M_2, \dots, M_R\}$, 感兴趣的参数 θ 对所有模型都有相同的解释意义。假设模型平均估计 $\hat{\theta} = \sum_i \omega_i \hat{\theta}_i$ 近似服从正态分布, 其中 ω_i 和 $\hat{\theta}_i$ 分别对应模型 M_i 的权重和参数 θ 的估计。计算参数 θ 的 $100(1-\alpha)\%$ Wald 置信区间形式为

$$\hat{\theta} \pm z_\alpha \cdot \widehat{se}(\hat{\theta})$$

其中 z_α 是标准正态分布的 $100(1-\alpha)\%$ 分位数, $\widehat{se}(\hat{\theta})$ 是 $\hat{\theta}$ 方差的估计。尽管这种方法很简单, 但是计算一个合适的 $se(\hat{\theta})$ 的估计是很困难的。由 Burnham [5] 和 Anderson 等人提出的两种 $\widehat{se}(\hat{\theta})$ 估计为

$$\widehat{se}(\hat{\theta}) = \sum_i \omega_i \sqrt{\left(\widehat{se}(\hat{\theta}) \cdot t_{v_i} / z_\alpha\right)^2 + \left(\hat{\theta}_i - \hat{\theta}\right)^2} \quad (2.1)$$

$$\widehat{se}(\hat{\theta}) = \sum_i \omega_i \sqrt{\left[\left(\widehat{se}(\hat{\theta}) \cdot t_{v_i} / z_\alpha\right)^2 + \left(\hat{\theta}_i - \hat{\theta}\right)^2\right]} \quad (2.2)$$

t_{v_i} 是自由度为 v_i 的 t 分布的分位数, v_i 是与模型 M_i 相关的残差的自由度。由 Cauchy-Schwarz 不等式可知, $\widehat{se}_2 \geq \widehat{se}_1$, 则根据(2.2)式得到的 Wald 置信区间比由(2.1)式得到的 Wald 置信区间长度宽, 故其覆盖率大, 本文我们只考虑(2.2)式得到的 Wald 置信区间。Wald 区间在覆盖率方面表现得很差, 并且忽视了 $\hat{\theta}$ 的非正态性[8], 下面基于单个模型来获取参数 θ 的模型平均置信区间。

假设恰有一个候选模型是真实的, 记 c_i 为未知的指示变量, 如果模型 M_i 为真, 则 $c_i = 1$ 否则 $c_i = 0$, 假设我们知道 $\hat{\theta}_i$ 的真实分布记为 $G_i(\cdot)$, 参数 θ 的置信区间 $[\theta_L, \theta_U]$ 满足下列方程

$$\begin{cases} \sum_{i=1}^R c_i G_i(\theta_L) = \alpha \\ \sum_{i=1}^R c_i G_i(\theta_U) = 1 - \alpha \end{cases} \quad (2.3)$$

因为 c_i 是未知的, 我们用与模型相对应的基于某种信息准则的模型权重 ω_i 来估计(2.3)中的 c_i , 则参数 θ 的置信区间 $[\theta_L, \theta_U]$ 满足下列方程

$$\begin{cases} \sum_{i=1}^R \omega_i G_i(\theta_L) = \alpha \\ \sum_{i=1}^R \omega_i G_i(\theta_U) = 1 - \alpha \end{cases} \quad (2.4)$$

如果不知道 $\hat{\theta}_i$ 的真实分布, 假设每个模型的学生化的参数估计 $\frac{\hat{\theta}_i - \theta}{\widehat{se}(\hat{\theta}_i)}$ 服从标准正态分布或 t 分布,

即 $G_i(\cdot)$ 为标准正态分布或者 t 分布的分布函数, 获取的参数 θ 的 $100(1-\alpha)\%$ 的置信区间即为 z 版本或者 t 版本的 MATA 区间, 本文我们考虑 t 版本的 MATA 区间[8]。

但是如果相应变量是偏态的, 或者正态假设不明显时, 使用 Wald 区间和 MATA 区间会导致比较差的覆盖率。本文利用参数 Bootstrap 方法构造变量的分布, 即用参数估计 $\hat{\theta}_i$ 的 Bootstrap 分布 $\widehat{G}_i(\cdot)$ 来近似 (2.4) 中的分布 $G_i(\cdot)$, 从而计算参数的模型平均置信区间。

设样本 $X = (X_1, X_2, \dots, X_n)$ 来自模型 M_i 的分布 $F_i(X, \theta)$ 容量为 n 的样本, $\hat{\theta}_i$ 是参数 θ 的极大似然估计。下面给出计算参数模型平均估计的计算步骤。

- 对于来自分布为 $F_i(X, \theta)$ 的样本, 选取 Bootstrap 样本量 B , 比如 $B = 2000$;
- 对于 $i = 1, \dots, B$, 从 $F_i(X, \hat{\theta}_i)$ 产生容量为 n 的 Bootstrap 样本;
- 计算出 θ_i 的极大似然估计 $\hat{\theta}_i^{(j)}$;
- 由 $(\hat{\theta}_i^{(1)}, \dots, \hat{\theta}_i^{(B)})$ 给出 $\hat{\theta}_i$ 的 Bootstrap 分布 $\widehat{G}_i(\cdot)$ 。

求 θ_L 和 θ_U 满足下列方程

$$\begin{cases} \sum_{i=1}^R \omega_i \widehat{G}_i(\theta_L) = \alpha/2 \\ \sum_{i=1}^R \omega_i \widehat{G}_i(\theta_U) = 1 - \alpha/2 \end{cases} \quad (2.5)$$

其中 ω_i 为某种信息准则下模型 M_i 的权重。那么 $[\theta_L, \theta_U]$ 为参数 θ 的 Bootstrap 模型平均置信区间。

3. 数据模拟

本节对 Bootstrap 模型平均置信区间的优良性进行评判, 并与已有的与 Wald 区间, MATA 区间进行比较, 最终从区间的频率性质和区间长度这两方面进行评价。区间的频率性质用以下两个指标: 覆盖率, 上下错误率。

进行模型选择时, 权重的选择是基于信息准则 AIC

$$\omega_i = \frac{\exp(-\text{AIC}_i/2)}{\sum_{j=1}^R \exp(-\text{AIC}_j/2)}$$

其中 AIC_i 为第 i 个模型的 AIC。

3.1. 线性模型

考虑正态线性模型, 包括两个预测变量 x_1 和 x_2 , 5 个候选模型 $\{M_1, M_2, \dots, M_5\}$ 。所有模型的响应变量 $y_i \sim N(\mu_i, \sigma^2)$, 并且 μ_i 指定如下:

$$M_1: \mu_i = \beta_0$$

$$M_2: \mu_i = \beta_0 + \beta_1 x_{1,i}$$

$$M_3: \mu_i = \beta_0 + \beta_2 x_{2,i}$$

$$M_4: \mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$M_5: \mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_{12} x_{1,i} x_{2,i}$$

随意设置 $\sigma = 1$, 选择 $\beta_0 = 1, \beta_1 = \beta_2 = 0.3, \beta_{12} = 0.1$, 这样所有的模型都倾向于依据样本容量的范围分配非平凡权重。 x_1 和 x_2 是在每轮模拟中随机产生的, 并且 $X_{1,i} \sim N(0, 1), X_{2,i} \sim \text{Gamma}(\alpha = 2, \beta = 1)$, 样本量大小分别取 $n = 20, 25, \dots, 100$ 这 17 种情形(见表 1)。对于感兴趣的参数选取为预测空间 (x_1, x_2) 各个点的 μ 值, 为了简单起见, 这里只考虑 x_1 生成分布的 50% 处分位数和 x_2 生成分布的 90% 处分位数。这里模

拟次数为 10^5 次。在计算 Bootstrap 模型平均置信区间时, Bootstrap 样本量取 $B = 2000$ 。

图 1 展示了响应变量均值的三种区间的上下错误率和覆盖率随样本容量 n 的变化形势, 图 1 的左图上方为三种区间的上错误率。为了表示方便, 下错误率取负值后展示在图 1 的左图下方。表 1 展示了 3 种区间估计的平均区间长度。由图 1 和表 1 可以看出:

- 1) 下错误率方面, Bootstrap 区间要好于其它两种区间, 更接近名义错误率。上错率和覆盖率方面, Bootstrap 区间整体要好于 Wald 区间, 略逊于 MATA 区间。
- 2) 平均区间长度方面, Bootstrap 区间是最窄的。
- 3) 随着样本量的增加, 三种区间估计方法都逐渐接近名义覆盖率, 并且平均区间长度区别越来越小。

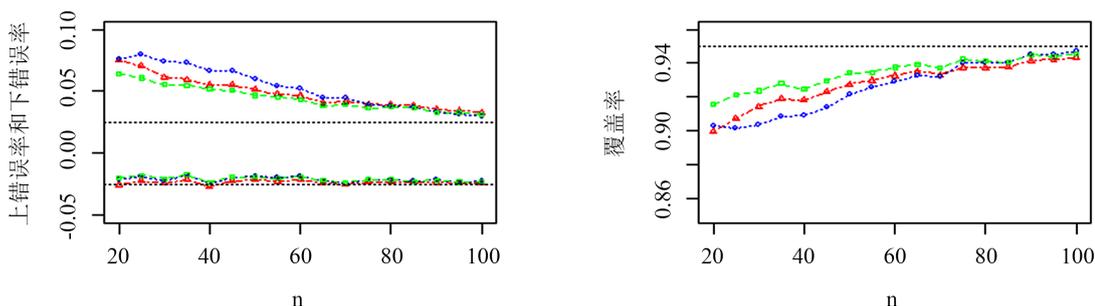


Figure 1. Error rate and coverage of Bootstrap (red triangle), MATA (green square) and Wald (blue circle) interval estimates of linear model mean

图 1. 线性模型均值的 Bootstrap (红色三角)、MATA (绿色方形)、Wald (蓝色圆形) 区间估计的上下错误率、覆盖率

Table 1. Average interval lengths of three interval estimates of linear model mean

表 1. 线性模型均值的三种区间估计的平均区间长度

n	Bootstrap	Wald	MATA
20	1.776	1.885	1.873
25	1.569	1.649	1.636
30	1.422	1.484	1.471
35	1.309	1.359	1.348
40	1.216	1.258	1.249
45	1.139	1.174	1.166
50	1.074	1.103	1.097
55	1.019	1.044	1.039
60	0.967	0.988	0.985
65	0.925	0.943	0.941
70	0.886	0.902	0.900
75	0.852	0.866	0.865
80	0.820	0.832	0.832
85	0.791	0.802	0.802
90	0.765	0.775	0.775
95	0.742	0.751	0.751
100	0.720	0.727	0.727

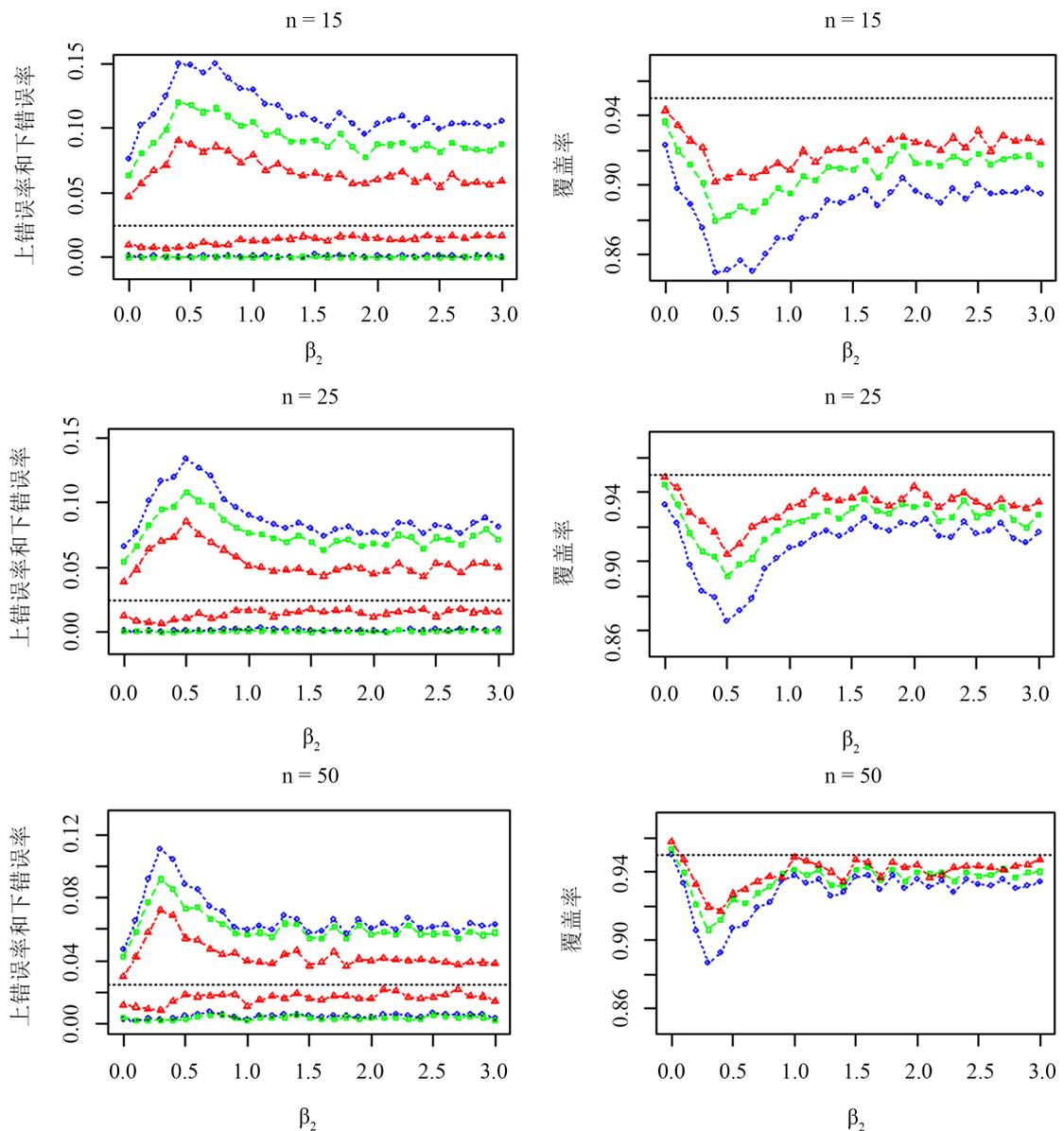


Figure 2. Upper and lower error rates and coverage of Bootstrap (red triangle), MATA (green square) and Wald (blue circle) interval estimates of the mean of lognormal model

图 2. 对数正态模型均值的 Bootstrap (红色三角)、MATA (绿色方形)、Wald (蓝色圆形) 区间估计的上下错误率、覆盖率

Table 2. Average interval lengths of three interval estimates of lognormal model when $n = 50$

表 2. $n = 50$ 时对数正态模型的三种区间估计的平均区间长度

β_2	Wald	MATA	Bootstrap
0	0.929	0.953	0.947
0.1	1.019	1.046	1.04
0.2	1.152	1.178	1.175
0.3	1.307	1.333	1.332
0.4	1.495	1.52	1.524

Continued

0.5	1.689	1.717	1.72
0.6	1.878	1.912	1.913
0.7	2.069	2.112	2.109
0.8	2.274	2.327	2.317
0.9	2.503	2.566	2.554
1	2.772	2.843	2.826
1.1	3.047	3.125	3.104
1.2	3.377	3.465	3.443
1.3	3.73	3.826	3.802
1.4	4.131	4.238	4.216
1.5	4.557	4.675	4.645
1.6	5.026	5.156	5.123
1.7	5.546	5.689	5.654
1.8	6.145	6.303	6.266
1.9	6.806	6.982	6.939
2	7.529	7.723	7.68
2.1	8.343	8.559	8.515
2.2	9.173	9.41	9.355
2.3	10.1	10.361	10.303
2.4	11.202	11.492	11.424
2.5	12.403	12.724	12.665
2.6	13.652	14.005	13.922
2.7	15.222	15.615	15.532
2.8	16.669	17.1	17.014
2.9	18.418	18.894	18.794
3	20.331	20.857	20.75

3.2. 对数正态模型

下面考虑对数正态的偏态数据情形。设观察值 Y_{ij} 满足 $\log(Y_{ij}) \sim N(\beta_i, \sigma^2)$, 我们去估计 Y_{ij} 期望 $E(Y_{ij}) \equiv \mu_i = \exp\left(\beta_i + \frac{\sigma^2}{2}\right)$ ($i=1, 2; j=1, \dots, n$)。模型平均基于以下两个模型:

$$M_1: \beta_1 = \beta_2$$

$$M_2: \beta_1, \beta_2 \text{ 未指定}$$

我们设置 $\sigma^2 = 1$, $\beta_1 = 0$ 。我们分别考虑了 $\beta_2 = 0, 0.1, 0.2, \dots, 3$ 及样本容量 $n = 15, 25, 50$ 的不同情形, 模型数据由 M_1 产生。这里模拟次数为 5000 次。在计算 Bootstrap 模型平均置信区间时, Bootstrap 样本量取 $B = 2000$ 。

图 2 展示了 $n = 15, 25, 50$ 三种情形下均值的三种区间的上下错误率和覆盖率随 β_2 的变化形势, 图 2 的左图上方为三种区间的上错误率。为了表示方便, 下错误率取负值后展示在图 1 的左图下方。表 2 给

出了样本量 $n = 50$ 时三种区间的平均区间长度。由图 2 和表 2 可以看出

1) 当样本容量较小时, 比如 $n = 15$, Bootstrap 区间的上、下错误率和覆盖率要明显优于 MATA 区间和 Wald 区间, 并且更接近名义上、下错误率 2.5% 和名义置信水平 95%。随着样本容量的增大, 三种区间的表现性能越来越接近, Bootstrap 区间的上、下错误率和覆盖率总是优于 MATA 区间和 Wald₂ 区间, 并且更接近名义水平。

2) 当 β_2 逐渐增大时, 也就是两个模型 M_1 和 M_2 之间的差异越来越大时, 三种模型平均置信区间的表现越来越好。

3) 在样本量 $n = 50$ 时, Wald 区间长度最短。Bootstrap 区间的平均区间长度略低于 MATA 区间, 并且随着 β_2 的增大, 两种区间的平均区间长度的相对差异也很小。

4. 结论

无论是 Wald 区间还是 MATA 区间, 都是在参数估计或者参数估计的变换近似服从正态分布的假设下构建的, 本文提出了 Bootstrap 模型平均置信区间, 它是基于将参数估计的 Bootstrap 分布来近似其真实分布从而得到的一种模型平均置信区间。

当响应变量是正态数据时, 我们研究了在线性正态模型下 Wald 区间、MATA 区间和 Bootstrap 区间的表现性能。模拟结果表明, Bootstrap 数区间在下错误率方面优势明显, 一直稳定在名义下错误率附近。上错误率方面, Bootstrap 区间整体上也是仅次于 MATA 区间, 并且在平均区间长度方面, Bootstrap 区间较其他三种区间是最窄的。可以看出数 Bootstrap 区间缩短了区间长度, 牺牲了一些覆盖率的表现性能。

当响应变量是偏态数据时, 我们研究了对数正态模型下三种区间的表现性能。模拟结果表明, Bootstrap 区间要优于其他两种置信区间。特别是在上错误率方面, Bootstrap 区间提供了更低的更接近名义水平的上错误率。除此之外, 下错误率和覆盖率也优于其他两种区间。在样本量很小时, Bootstrap 区间的优势最明显。随着样本容量的增加, 三种区间的表现越来越接近。在平均区间长度方面, MATA 区间和 Bootstrap 区间相差不大。

因此, 当响应变量是偏态的, 甚至正态假设不明显时, 都会导致 MATA 区间或者 Wald 区间的错误率和覆盖率很差。Bootstrap 区间适应性比其它两个区间要强。特别是在响应变量是偏态的情况下, 覆盖率是最好的。所以在不知道响应变量是否是正态的情况下, Bootstrap 区间是一个不错的选择。

参考文献

- [1] Chatfield, C. (1995) Model Uncertainty, Data Mining and Statistical Inference. *Journal of the Royal Statistical Society: Series A*, **158**, 419-466. <https://doi.org/10.2307/2983440>
- [2] Wasserman, L. (2000) Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology*, **44**, 92-107. <https://doi.org/10.1006/jmps.1999.1278>
- [3] Wang, H. and Zhou, S.Z. (2013) Interval Estimation by Frequentist Model Averaging. *Communications in Statistics*, **42**, 4342-4356. <https://doi.org/10.1080/03610926.2011.647218>
- [4] Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model Selection: an Integral Part of Inference. *Biometrics*, **53**, 603-618. <https://doi.org/10.2307/2533961>
- [5] Hjort, N.L. and Claeskens, G. (2003) Frequentist Model Average Estimators. *Journal of the American Statistical Association*, **98**, 879-899. <https://doi.org/10.1198/016214503000000828>
- [6] Burnham, K. and Anderson, D. (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd Edition, Springer, New York.
- [7] Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model Selection: An Integral Part of Inference. *Biometrics*, **53**, 603-618. <https://doi.org/10.2307/2533961>
- [8] Turek, D. and Fletcher, D. (2012) Model-Averaged Wald Confidence Intervals. *Computational Statistics and Data Analysis*, **56**, 2809-2815. <https://doi.org/10.1016/j.csda.2012.03.002>

- [9] Fletcher, D. and Turek, D. (2012) Model-Averaged Profile Likelihood Intervals. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 38-51. <https://doi.org/10.1007/s13253-011-0064-8>
- [10] Kabaila, P., Welsh, A. and Abeysekera, W. (2016) Model-Averaged Confidence Intervals. *Scandinavian Journal of Statistics*, **43**, 35-48. <https://doi.org/10.1111/sjos.12163>
- [11] Kabaila, P. (2018) On the Minimum Coverage Probability of Model Averaged Tail Area Confidence Intervals. *Canadian Journal of Statistics*, **46**, 279-297. <https://doi.org/10.1002/cjs.11349>
- [12] Kabaila, P., Welsh, A.H. and Mainzer, R. (2017) The Performance of Model Averaged Tail Area Confidence Intervals. *Communications in Statistics-Theory and Methods*, **46**, 10718-10732. <https://doi.org/10.1080/03610926.2016.1242741>
- [13] Yu, W., Xu, W. and Zhu, L. (2014) Transformation-Based Model-Averaged Tail Area Inference. *Computational Statistics*, **29**, 1713-1726. <https://doi.org/10.1007/s00180-014-0514-1>