

大样本Gamma回归的最优子抽样

尚建敏, 张 齐

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2022年3月12日; 录用日期: 2022年4月5日; 发布日期: 2022年4月14日

摘 要

随着计算机行业的迅猛发展, 人类社会逐渐迈入大数据时代。面对大规模右偏性和厚尾分布的数据, Gamma回归模型发挥着非常重要的作用。然而如何快速并准确估计出Gamma回归中感兴趣参数成为值得思考的热点问题。在本文中, 我们提出两种两步算法分别有效地逼近 ϕ 已知Gamma回归和 ϕ 未知Gamma回归在全数据下的最大似然估计, 从而解决了单参数与双参数大样本Gamma回归估计问题。首先在 ϕ 已知情况下, 可证明出在给定全数据下一般子抽样估计量渐近服从正态分布, 推导出使估计量渐近均方差最小的最优子抽样概率。为了进一步降低计算量, 我们还提出了另一种最优子抽样概率。由于最优子抽样概率取决于未知参数, 我们还提出了单参数两步算法。其次在 ϕ 未知情况下, 我们基于单参数两步算法提出了双参数两步算法。最后使用数值模拟表明两种算法的计算效率高, 也证实了通过单参数两步算法得到的估计量与双参数两步算法得到的估计量差距不明显。

关键词

大样本, Gamma回归, 最优子抽样, 抽样概率

Optimal Subsampling for Large-Sample Gamma Regression

Jianmin Shang, Qi Zhang

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Mar. 12th, 2022; accepted: Apr. 5th, 2022; published: Apr. 14th, 2022

Abstract

With the rapid development of computer industry, human society is gradually moving into the era of big data. Gamma regression models play a very important role in the face of large-sample right-skewed and thick-tailed data. However, how to quickly and accurately estimate the parame-

ters of interest in Gamma regression has become a hot issue to be considered. In this paper, we propose two two-step algorithms to efficiently approximate the maximum likelihood estimates of ϕ known Gamma regression and ϕ unknown Gamma regression under full data, respectively, thus solving the single-parameter and two-parameter large-sample Gamma regression estimation problems. Firstly, in the case where ϕ is known, it can be shown that the general subsampling estimates asymptotically obey a normal distribution given the full data, and the optimal subsampling probability that minimizes the asymptotic mean square error of the estimates is derived. To further reduce the computational effort, we also propose an alternative optimal subsampling probability. Since the optimal subsampling probability depends on the unknown parameters, we also propose a single-parameter two-step algorithm. Secondly, in the case of ϕ unknown, we propose a two-parameter two-step algorithm based on the one-parameter two-step algorithm. Finally, using numerical simulations, we show that the two algorithms are computationally efficient and also confirm that the difference between the estimates obtained by the one-parameter two-step algorithm and those obtained by the two-parameter two-step algorithm is not significant.

Keywords

Large Sample, Gamma Regression, Optimal Subsampling, Sampling Probability

Copyright © 2022 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

随着大数据兴起, 科学家们面临的数据规模也逐渐庞大, 前所未有的数据量为研究人员提供了前所未有的机遇和挑战。然而如何在有限的计算资源下处理规模巨大、结构复杂的数据越来越受到人们的关注。

针对线性回归模型, 目前处理大规模数据比较流行的方法有分治方法(Lin 等, 2011) [1]、数据流在线更新方法(Schifano 等, 2016) [2]以及子抽样方法(Politis 等, 1999) [3]。Ma 和 Sun (2015) [4]在子抽样方法(Politis 等) [3]基础上提出了杠杆抽样方法, 其方法成功的关键是利用不同权重构造非均匀抽样概率, 使有影响的数据点以高概率抽样(Ma, Yu 和 Mahoney, 2015) [5]。Ma 等(2020) [6]还研究了杠杆抽样估计的渐近正态性和渐近无偏性。还有很多将子抽样方法应用到其它问题上, Drineas 等[7]针对 L_2 回归模型提出了子抽样算法。Drineas 等(2011) [8]提出两种子抽样算法来解决线性回归中最小二乘逼近问题。Dhillon 等(2013) [9]针对线性回归中的普通最小二乘快速估计问题使用了子抽样算法。最优子抽样方法是子抽样方法的延伸之一, Zhu 等(2015) [10]针对线性回归模型提出了两种子抽样方法, 即最优子抽样方法和预测长度的子抽样方法。Wang 等[11]针对线性回归子抽样提出基于信息的最优子抽样方法(IBOSS)。

在广义线性模型下, Chen 和 Xie [12]在广义线性模型(GLM)上利用分治方法处理海量数据。Wang 等(2018) [13]在 logistic 回归模型上提出最优子抽样方法来有效地逼近在全数据下的最大似然估计, 该方法通过最小化估计量的渐近均方误差来确定最优子抽样概率。Yao 等(2018) [14]在 Wang 等[13]基础上研究了多分类 logistics 回归模型的最优子抽样算法。Wang 和 Ma (2021) [15]研究了分位数回归的最优子抽样, 推导出一般子抽样估计量的渐近性质以及两种最优子抽样概率。此外, 李莉莉等(2022) [16]针对岭回归模型提出最优子抽样算法, 从而处理了变量间存在多重共线性问题, 并证明了估计量的渐近性质。牛晓阳等(2022) [17]对非参数局部多项式回归模型提出最优子抽样算法, 提出 OPT 和 PL 两种抽样算法。

Gamma 回归模型作为广义线性模型中的一个特例[18], 它在很多方面有所应用, 比如周围等(2021) [19]将 Gamma 分布应用在海面目标检测中, 并介绍了在 Gamma 分布杂波背景下的检测器; 吴林林(2014) [20]利用 Gamma 回归研究了雨滴谱对移动双偏振雷达的影响; 王俊轶(2019) [21]利用 Gamma 回归研究了图像质量的影响因素。针对 Gamma 回归估计问题, 陈超(2015) [22]总结了双参数 Gamma 分布参数估计方法, 同时利用矩思想, 证明了形状参数存在的相合性和渐近正态性。针对双参数 Gamma 回归多重共线性问题, Muhammad Qasim 等(2018) [23]使用了 Liu 估计器。根据模拟研究, 表明 Liu 估计器总是优于最大似然估计。左卫兵和钱莉等[24]针对 Gamma 回归模型的多重共线性问题, 提出了 Gamma 回归模型的几乎无偏岭估计方法, 同时给出了该估计优于极大似然估计和一般岭估计的存在性定理。

在本文中, 我们将最优子抽样方法应用到 Gamma 回归模型中。本文的剩余部分组织如下: 在第二部分中, 我们分别建立了单参数 Gamma 回归和双参数 Gamma 回归的最大似然估计。在第三部分中, 我们首先考虑了在 ϕ 已知情况下 Gamma 回归的一般子抽样算法。通过三条正则性假设, 在给定全数据下, 我们建立了一般子抽样算法估计量的相合性和渐近正态性。最后我们考虑了在 ϕ 未知情况下 Gamma 回归的一般子抽样算法。在第四部分中, 我们基于 ϕ 已知情况下一般子抽样估计量的渐进性质, 提出两种最优子抽样概率。由于最优子抽样概率取决于未知量, 我们提出了单参数 Gamma 回归的两步算法来近似逼近全数据下的估计量。在 ϕ 未知情况下, 我们提出了双参数 Gamma 回归的两步算法。在第五部分中, 我们进行了广泛的仿真研究, 以验证我们的方法的有效性。在第六部分中, 我们总结了本文。所有证明细节见附录。

2. Gamma 回归

假设 $\{(x_i, y_i)\}_{i=1}^N$ 是一个独立且同分布的随机变量序列, 当响应变量 y_i 为正偏态且服从伽马分布时, Gamma 回归模型非常有用。假设在给定协变量 $x_i \in R^d$ 条件下, 响应变量 $y_i \in R$ 服从伽马分布, 则下述模型

$$\begin{cases} f(y_i; \mu_i, \phi) = \exp\left(\frac{y_i/\mu_i - \ln(1/\mu_i)}{-\phi} + \frac{1-\phi}{\phi} \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right)\right), & y_i > 0. \\ g(\mu_i) = x_i^T \beta. \end{cases} \quad (1)$$

称为 Gamma 回归模型。其中, μ_i 是响应变量 y_i 的数学期望且有限, 即 $E(y_i) = \mu_i$; 假设 $\beta = (\beta_1, \beta_2, \dots, \beta_d)^T$ 是 d 维感兴趣参数, 属于 R^d 的紧子集; $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ 为 Gamma 函数; $g(\cdot)$ 是连接函数, 它建立了样本总体均值与线性预测值之间的关系, 在 Gamma 回归模型中我们一般采用对数连接函数, 即 $g(\mu_i) = \ln(\mu_i) = x_i^T \beta$, 从而 $\mu_i = \exp(x_i^T \beta)$; ϕ 为离散参数。下面我们分别在 ϕ 已知和 ϕ 未知这两种情况下估计我们所感兴趣的参数 β 。

2.1. ϕ 已知的 Gamma 回归

我们先考虑 ϕ 已知情况。由于 ϕ 已知, 根据(1)式我们很容易可以写出 Gamma 回归模型, 即

$$f(y_i | x_i; \beta) = \exp\left(\frac{y_i/\exp(x_i^T \beta) + x_i^T \beta}{-\phi} + \frac{1-\phi}{\phi} \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right)\right), \quad y_i > 0.$$

Gamma 回归模型是广义线性模型的一种特殊情况。在 Y 关于 β 的最大化对数似然函数的运算过程中, 可以只考虑含参数 β 的项, 则 Y 的对数似然函数可以简洁的表达为

$$l(\beta) = \sum_{i=1}^N \left[\frac{y_i/\exp(x_i^T \beta) + x_i^T \beta}{-\phi} \right]. \quad (2)$$

估计 Gamma 回归中未知参数 β 通常是最大化对数似然函数。然而, $\hat{\beta}_{MLE}$ 一般没有封闭解, 通常采用迭

代方法来进行数值求解, 牛顿迭代法是一种常用的迭代方法。根据式(2), 我们可得到下面公式,

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \left[\sum_{i=1}^N \frac{y_i x_i x_i^T}{\exp(x_i^T \hat{\beta}^{(t)})} \right]^{-1} \sum_{i=1}^N \frac{y_i x_i - \exp(x_i^T \hat{\beta}^{(t)}) x_i}{\exp(x_i^T \hat{\beta}^{(t)})}. \quad (3)$$

2.2. ϕ 未知的 Gamma 回归

现在我们考虑 ϕ 未知情况。根据(1)式我们可以写出, 在已知协变量 x_i 情况下, 响应变量 y_i 的概率密度函数为

$$f(y_i | x_i; \beta, \phi) = \exp \left[\frac{y_i / \exp(x_i^T \beta) + x_i^T \beta}{-\phi} + \frac{1-\phi}{\phi} \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma \left(\frac{1}{\phi} \right) \right], \quad y_i > 0.$$

由于 ϕ 未知, 所以回归中有两个未知参数 β 与 ϕ 待估计, 记 $\theta = (\beta, \phi)^T$ 。估计 θ 通常是最大化对数似然函数, 然而, $\hat{\theta}_{MLE}$ 一般没有封闭解, 所以我们通常使用牛顿迭代算法。 Y 关于参数 θ 的对数似然函数可以表达为

$$l(\theta) = \sum_{i=1}^N \left[\frac{y_i / \exp(x_i^T \beta) + x_i^T \beta}{-\phi} + \frac{1-\phi}{\phi} \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma \left(\frac{1}{\phi} \right) \right].$$

根据牛顿迭代公式, 我们可得到下面公式,

$$\hat{\theta}^{(t+1)} = \hat{\theta}^{(t)} + \left[I(\hat{\theta}^{(t)}) \right]^{-1} s(\hat{\theta}^{(t)}). \quad (4)$$

其中, $S(\theta)$ 、 $I(\theta)$ 分别是 Gamma 回归模型的得分函数和 Fisher 信息矩阵。

在(3)式中, 我们发现每次迭代需要 $O(Nd^2)$ 时间, ζ 为所需迭代次数, 因此, 整个优化过程需要 $O(\zeta Nd^2)$ 时间。由于 θ 是 $d+1$ 维向量, 所以(4)式每次迭代需要 $O(N(d+1)^2)$ 时间, 整个优化过程需要 $O(\zeta N(d+1)^2)$ 时间。然而对于超大样本, (3)式单次运行的所需计算时间太长了, 更别说计算(4)式了。所以我们使用子抽样算法来解决超大样本问题。

3. 一般子抽样

对于海量数据, 为了缩短其计算时间, 我们利用抽样概率 $\pi_i (i=1, 2, \dots, N)$, 从全部数据中有放回随机抽取 $n (\ll N)$ 个子样本, 其中 $\sum_{i=1}^N \pi_i = 1$, π_i 既取决于完整数据 $D_N = \{(x_i, y_i)\}_{i=1}^N$, 也取决于响应变量。

我们需要使用 π_i 的倒数作为子样本对数似然函数的权重, 否则, 估计结果往往是有偏差的。我们首先考虑在 ϕ 已知情况下的 Gamma 回归模型的子抽样。

3.1. ϕ 已知的 Gamma 回归的子抽样

子样本表示为 (x_i^*, y_i^*) , 子抽样概率是 $\pi_i^* (i=1, 2, \dots, n)$, $\tilde{\beta}$ 为子样本估计器, 通过最大化(5)式的加权对数似然函数, 可得到基于子样本的估计值。

$$l^*(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\pi_i^*} \left[\frac{y_i^* / \exp((x_i^*)^T \beta) + (x_i^*)^T \beta}{\phi} \right]. \quad (5)$$

最大化(5)式以获得基于子样本的估计值 $\tilde{\beta}$ 。由于 $l^*(\beta)$ 的凸性, 最大化可以用牛顿迭代法来实现, 反

复应用以下公式, 直到 $\tilde{\beta}^{(t+1)}$ 和 $\tilde{\beta}^{(t)}$ 足够接近,

$$\tilde{\beta}^{(t+1)} = \tilde{\beta}^{(t)} + \left[\sum_{i=1}^n \frac{y_i^* x_i^* (x_i^*)^T}{\pi_i^* \exp((x_i^*)^T \tilde{\beta}^{(t)})} \right]^{-1} \sum_{i=1}^n \frac{y_i^* x_i^* - \exp((x_i^*)^T \tilde{\beta}^{(t)}) x_i^*}{\pi_i^* \exp((x_i^*)^T \tilde{\beta}^{(t)})}. \quad (6)$$

在 ϕ 已知情况下, 我们研究子样本估计量 $\tilde{\beta}$ 的渐近性质。在整篇论文中, $O_p(1)$ 指依概率有界。 $\|v\|$ 表示向量 v 的欧几里德范数, 即 $\|v\| = \langle v^T v \rangle^{1/2}$ 。下面我们列出一些正则性假设。

假设 1. 当 $N \rightarrow \infty$ 时, 有 $G_N = \frac{1}{N} \sum_{i=1}^N \frac{y_i x_i x_i^T}{\phi \exp(x_i^T \hat{\beta}_{MLE})}$ 依概率收敛于正定矩阵, 并且

$$\frac{1}{N} \sum_{i=1}^N \left[\frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} \right]^4 = O_p(1).$$

假设 2. $\frac{1}{N^3} \sum_{i=1}^N \frac{\|x_i\|^8}{\pi_i^2} = O_p(1)$ 。如果 $\rho \geq 0$, 有 $\frac{1}{N} \sum_{i=1}^N \left[\frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right]^{4+2\rho} = O_p(1)$,

$$\frac{1}{N^{3+2\rho}} \sum_{i=1}^N \pi_i^{-2-2\rho} \|x_i\|^{4+2\rho} = O_p(1).$$

假设 3. $\frac{1}{N^2} \sum_{i=1}^N \left[\frac{y_i / \exp(x_i^T \beta) + x_i^T \beta}{\pi_i} \right]^2 = O_p(1)$, $\frac{1}{N} \sum_{i=1}^N \frac{y_i}{\exp(x_i^T \beta)} \|x_i\|^3 = O_p(1)$ 。

假设 1 中 $l(\beta)$ 关于 β 的二阶导函数是渐近非奇异的, 而且 $y_i / \exp(x_i^T \hat{\beta}_{MLE})$ 的四阶矩是依概率有界的, 这个假设是很易实现的。在假设 2 中, 当 $\rho \geq 0$ 时, $y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1$ 的四阶矩是依概率有界的, 同时, 我们还对子抽样概率引入了一些条件, 如果子抽样概率选为均匀子抽样概率 $\pi_i = N^{-1}$, 我们不难发现, 协变量 x_i 的 $4+2\rho$ 阶矩依概率有界。在假设 1 和假设 2 成立的条件下, 假设 3 对指数分布族均成立, 这个假设是合理的。

定理 1. 如果假设 1、2 和 3 均成立, 当 $N \rightarrow \infty$ 和 $n \rightarrow \infty$ 时, 在给定全数据 D_N 的条件下, 近似误差 $\tilde{\beta} - \hat{\beta}_{MLE}$ 依概率收敛于 0, 同时收敛速度是 $n^{-1/2}$, 也就是说, 对任意 $\varepsilon > 0$, 存在一个有限的 Δ_ε 和 n_ε , 使得

$$P\left(\|\tilde{\beta} - \hat{\beta}_{MLE}\| \geq n^{-1/2} \Delta_\varepsilon \mid D_N\right) < \varepsilon.$$

对任意 $n > n_\varepsilon$ 均成立。

定理 2. 如果假设 1、2 和 3 均成立, 当 $N \rightarrow \infty$ 和 $n \rightarrow \infty$ 时, 在给定全数据 D_N 的条件下, 近似误差 $\tilde{\beta} - \hat{\beta}_{MLE}$ 的条件分布渐近服从正态分布, 即

$$V_N^{-1/2} (\tilde{\beta} - \hat{\beta}_{MLE}) \rightarrow N(0, I).$$

其中

$$V_N = G_N^{-1} V_n G_N^{-1}$$

$$V_n = \frac{1}{nN^2} \sum_{i=1}^N \frac{\left\{ \frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right\}^2 x_i x_i^T}{\phi^2 \pi_i}$$

定理 1 和定理 2 表明, 使用足够大的子样本, 可以使近似误差 $\tilde{\beta} - \hat{\beta}_{MLE}$ 尽可能小, 并且其渐近服从正态分布。在这里当 n 和 N 都趋于无穷, 对它们的相对阶数并没有限制。即使 n 大于 N , 这个定理仍然

有效, 但是过抽样在计算上没有好处, 所以在实践中 $n \ll N$ 是很典型的。下面我们考虑在 ϕ 未知情况下的 Gamma 回归模型的子抽样。

3.2. ϕ 未知的 Gamma 回归的子抽样

在 ϕ 未知下, 假设子样本为 (x_i^*, y_i^*) , 子抽样概率是 $\pi_i^* (i=1, 2, \dots, n)$, $\tilde{\theta}$ 为子样本估计器, 通过最大化(7)式的加权对数似然函数, 可得到基于子样本的估计值。

$$l^*(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\pi_i^*} \left[\frac{y_i^* / \exp\left((x_i^*)^T \beta\right) + (x_i^*)^T \beta}{\phi} + \frac{1-\phi}{\phi} \ln y_i^* - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right) \right]. \quad (7)$$

我们假设 δ_i 表示是否选择了第 i 个数据点的指标变量, 即当 (x_i, y_i) 在子样本时, $\delta_i = 1$, 否则 $\delta_i = 0$ 。我们假设 δ_i 的分布是带参数的伯努利分布, 即

$$P(\delta_i = 1 | x_i, y_i) = \pi_i, i = 1, 2, \dots, N.$$

(7)式可以等价地写成

$$l^*(\theta) = \frac{1}{n} \sum_{i=1}^N \frac{\delta_i}{\pi_i^*} \left[\frac{y_i / \exp(x_i^T \beta) + x_i^T \beta}{\phi} + \frac{1-\phi}{\phi} \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right) \right].$$

最大化(7)式以获得基于子样本的估计值 $\tilde{\theta}$ 。由于 $l^*(\theta)$ 的凸性, 最大化可以用牛顿迭代法来实现, 反复应用以下公式, 直到 $\tilde{\theta}^{(t+1)}$ 和 $\tilde{\theta}^{(t)}$ 足够接近,

$$\tilde{\theta}^{(t+1)} = \tilde{\theta}^{(t)} + \left[I^*(\tilde{\theta}^{(t)}) \right]^{-1} S^*(\tilde{\theta}^{(t)}). \quad (8)$$

其中 $S^*(\theta) I^*(\theta)$ 分别是(7)式的得分函数和 Fisher 信息矩阵。

4. 最优子抽样

我们发现无论在(6)式, 还是在(8)式, 最关键的是选择一个最佳的子抽样概率 π_i 。使用均匀子抽样概率 $\pi = \{\pi_i = N^{-1}\}_{i=1}^N$ 的算法往往不是“最优”的。从定理 2 可以看出, 渐近方差矩阵 V_N 取决于子抽样概率。因此, 我们可以导出使渐近方差矩阵最小的最优子抽样概率。由于 V_N 是矩阵, “最小化”含义需要定义, 这里我们采用 A-最优性和 L-最优性准则(Kiefer, 1959 [25])。在 ϕ 已知情况下, 通过采用两种不同的最优性准则, 下面的定理分别给出了两种最优子抽样概率, 同时这两种最优子抽样概率分别最小化了 $\tilde{\beta}$ 和 $G_N \tilde{\beta}$ 的渐近均方误差。

定理 3. 使迹 $tr(V_N)$ 最小的最优子抽样概率为

$$\pi_i^{Aopt} = \frac{\left| \left(y_i / \exp\left(x_i^T \hat{\beta}_{MLE}\right) \right) - 1 \right| \cdot \|G_N^{-1} x_i\|}{\sum_{j=1}^N \left| \left(y_j / \exp\left(x_j^T \hat{\beta}_{MLE}\right) \right) - 1 \right| \cdot \|G_N^{-1} x_j\|}, \quad i = 1, 2, \dots, N. \quad (9)$$

使迹 $tr(V_n)$ 最小的最优子抽样概率为

$$\pi_i^{Lopt} = \frac{\left| \left(y_i / \exp\left(x_i^T \hat{\beta}_{MLE}\right) \right) - 1 \right| \cdot \|x_i\|}{\sum_{j=1}^N \left| \left(y_j / \exp\left(x_j^T \hat{\beta}_{MLE}\right) \right) - 1 \right| \cdot \|x_j\|}, \quad i = 1, 2, \dots, N. \quad (10)$$

最优子抽样概率 π_i^{Aopt} 需要计算 $\|G_N^{-1} x_i\|$, 消耗 $O(Nd^2)$ 时间。而对于 π_i^{Lopt} 是需要计算 $\|x_i\|$, 其在计算上只需要消耗 $O(Nd)$ 时间, 所以 π_i^{Lopt} 在计算上优于 π_i^{Aopt} , 同时 π_i^{Lopt} 也避免对矩阵求逆和减少出现奇异

值的现象。要想使用 π_i^{Lopt} 和 π_i^{Aopt} ，我们发现(9)和(10)式的抽样概率均取决于 $\hat{\beta}_{MLE}$ ，这正是需要完整数据来估计参数所得到的，因此最优子抽样概率不能直接使用。为了更加清晰的表达，下面我们描述了在 ϕ 已知情况下 Gamma 回归估计的两步算法过程。

算法 1 单参数两步算法

- 第一步：根据均匀子抽样概率 $\pi = \{\pi_i = N^{-1}\}_{i=1}^N$ ，随机抽取大小为 n_0 的子样本，使用牛顿迭代方法，我们可以得到 β 一个先导估计值，即 $\tilde{\beta}_0 = \arg \max_{\beta} l^{*0}(\beta)$ ，其中 $l^{*0}(\beta)$ 表达式与(5)式相同。在(9)或(10)式中，我们将 $\hat{\beta}_{MLE}$ 替换为 $\tilde{\beta}_0$ ，从而获得与所选最优性准则相对应的近似最优子抽样概率 $\tilde{\pi}_i^{Aopt}$ 或者 $\tilde{\pi}_i^{Lopt}$ 。
- 第二步：用第一步中计算的 $\tilde{\pi}_i^{Aopt}$ 或 $\tilde{\pi}_i^{Lopt}$ 来随机抽取样本大小为 n 的子样本。结合这两个步骤中的 $n_0 + n$ 个子样本和其所对应的抽样概率，利用牛顿迭代法获得最终估计值 $\tilde{\beta}$ ，即

$$\tilde{\beta} = \arg \max_{\beta} \left\{ \frac{n_0}{n_0 + n} l^{*0}(\beta) + \frac{n}{n_0 + n} l^{*opt}(\beta) \right\}$$

其中， $l^{*opt}(\beta)$ 表达式与(5)式相同， π_i 由 $\tilde{\pi}_i^{Aopt}$ 或者 $\tilde{\pi}_i^{Lopt}$ 来代替。

在 ϕ 未知下，为了实现(8)式，关键还是在于如何选取子抽样概率。在这一节中，我们考虑使用(9)和(10)式的子抽样概率。由于 ϕ 未知，我们不难发现(9)式的子抽样概率是关于未知参数 ϕ 和 $\hat{\beta}_{MLE}$ 的，(10)式的子抽样概率取决于 $\hat{\beta}_{MLE}$ ，所以要想使用(9)或(10)式的子抽样概率，我们首先使用一部分样本将 ϕ 和 β 估计出来，即 $\tilde{\beta}_0$ 和 $\tilde{\phi}_0$ ，将 $\hat{\beta}_{MLE}$ 和 ϕ 分别替换为 $\tilde{\beta}_0$ 和 $\tilde{\phi}_0$ ，进而得到子抽样概率。具体算法如下。

算法 2 双参数两步算法

- 第一步：根据均匀子抽样概率 $\pi = \{\pi_i = N^{-1}\}_{i=1}^N$ ，随机抽取大小为 n_0 的子样本，使用牛顿迭代方法，我们可以得到 β 和 ϕ 的一个先导估计值，定义 $\tilde{\beta}_0$ 和 $\tilde{\phi}_0$ ，即 $\tilde{\theta}_0 = (\tilde{\beta}_0, \tilde{\phi}_0) = \arg \max_{\beta, \phi} l^{*0}(\theta)$ ，其中 $l^{*0}(\theta)$ 与(7)表达式相同。在(9)或(10)式中，我们将 $\hat{\beta}_{MLE}$ 替换为 $\tilde{\beta}_0$ ， ϕ 替换为 $\tilde{\phi}_0$ ，从而得到子抽样概率 $\tilde{\pi}_i^{Aopt}$ 或者 $\tilde{\pi}_i^{Lopt}$ 。
- 第二步：用第一步中计算的 $\tilde{\pi}_i^{Aopt}$ 或 $\tilde{\pi}_i^{Lopt}$ 来随机抽取样本大小为 n 的子样本。结合这两个步骤中的 $n_0 + n$ 个子样本和其所对应的抽样概率，利用牛顿迭代法获得最终估计值 $(\hat{\beta}, \hat{\phi})$ ，即

$$\hat{\theta} = (\hat{\beta}, \hat{\phi}) = \arg \max_{\beta, \phi} \left\{ \frac{n_0}{n_0 + n} l^{*0}(\theta) + \frac{n}{n_0 + n} l^{*opt}(\theta) \right\}$$

其中， $l^{*opt}(\theta)$ 具有与(7)相同的表达式， π_i 由 $\tilde{\pi}_i^{Aopt}$ 或者 $\tilde{\pi}_i^{Lopt}$ 来代替。

5. 模拟研究

在本节中，我们将使用第 4 节中提出的算法 1 以及算法 2 在 R 上进行数值模拟实验，分别评估其性能。假设 β 是 7 维向量，将参数 β 的真实值设为 $\beta_0 = (-0.5, -0.5, -0.5, -0.5, 0.5, 0.5, 0.5)^T$ 。我们考虑样本大小为 $N = 10000$ 的完整数据集，其中 X 分布情况如下，

- a) 协变量 X 服从均值为 0，方差为 Σ 的多元正态分布，即 $N(0, \Sigma)$ ，其中 Σ 是一个矩阵，所有对角元素等于 1，非对角元素等于 0.5；
- b) 协变量 X 服从均值非零，方差为 Σ 的多元正态分布，即 $N(1, \Sigma)$ ，其中 Σ 与 a) 相同；

- c) $X \sim 1/2N(1, \Sigma) + 1/2N(-1, \Sigma)$, 其中 Σ 与情况 a) 相同;
- d) 协变量 X 的分量是独立的, 且每个分量都服从指数分布, 速率参数为 2.

5.1. ϕ 已知的模拟研究

在 ϕ 已知情况下, 我们假设 $\phi = 2$, 在这一节中我们研究了 $\tilde{\beta}$ 的均方误差和 CPU 运行时间等, 来评估算法 1 的性能.

5.1.1. 数据集分布、子样本量与 $\tilde{\beta}$ 的 MSE

为了评估在算法 1 的性能, 我们对上述四种情况分别进行探讨, 研究数据集分布、子样本量与 MSE 的关系. 我们固定第一步子样本量 $n_0 = 200$, 第二步子样本大小 $n = 500, 1000, 1500, 2000, 2500, 3000$, 重复执行 $S = 1000$ 次, 每次循环要计算估计值 $\tilde{\beta}$ 的均方误差. 为了进行比较, 我们提供了子样本大小为 $n_0 + n$ 的均匀子抽样的均方误差结果, 并使用 1000 个 bootstrap 样本计算了全数据的估计值. 图 1 结果表明, 对于四种不同的分布, 算法 1 比均匀子抽样方法更有优势, 并且基于抽样概率 π_i^{Aopt} 的算法 1 优于基于 π_i^{Lopt} 的算法 1.

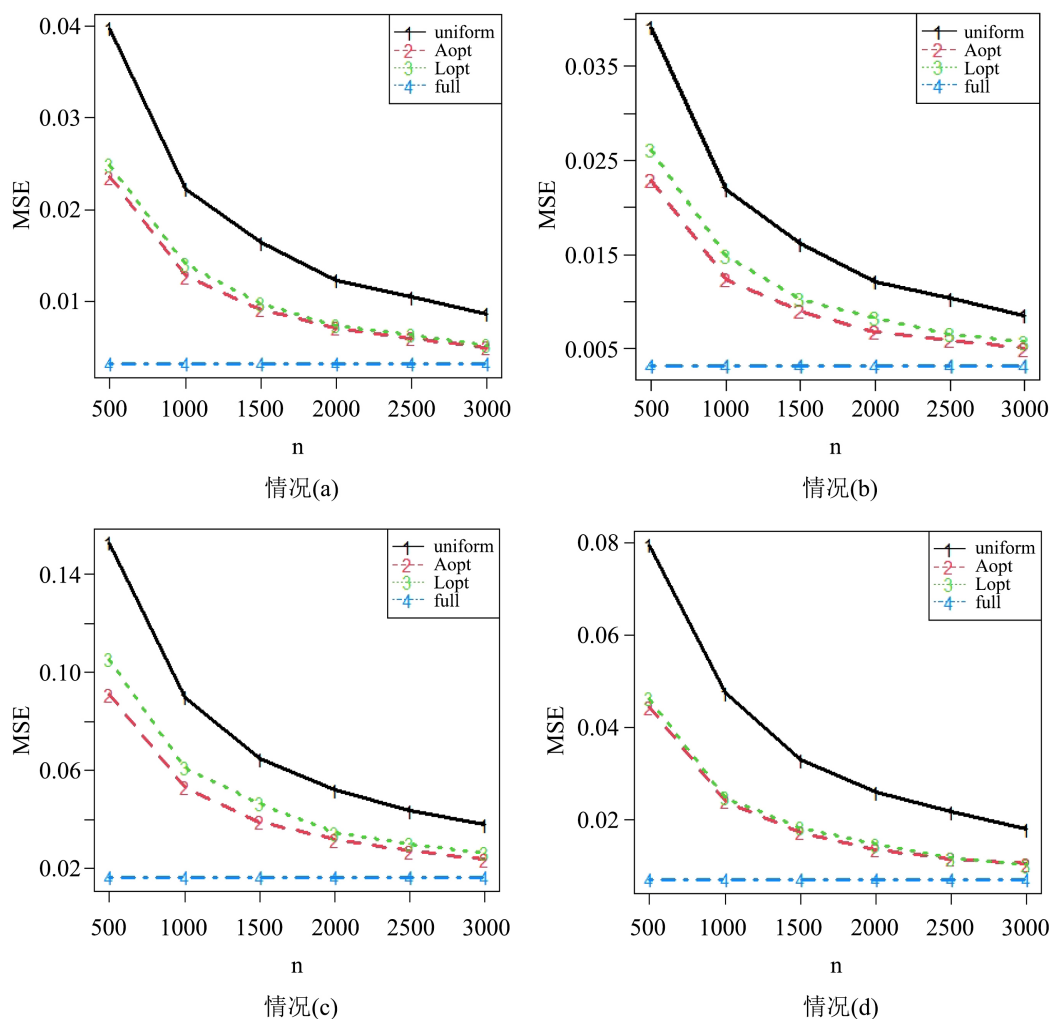


Figure 1. $n_0 = 200$, MSE of different data sets with n not fixed and different sampling methods

图 1. $n_0 = 200$, 不同数据集在 n 不固定和不同抽样方法下的 MSE

5.1.2. 子样本量与 CPU 运行时间

在第一步样本量 $n_0 = 200$ 和第二步子样本量为 $n = 100, 200, 300, 500, 700, 1000$ 下, 我们分别记录了 CPU 运行时间。经模拟研究, 我们在全数据下使用了 1000 次 bootstrap 抽样, 其 CPU 运行时间为 15.761 秒。表 1 给出了基于抽样概率为 π_i^{Aopt} 和 π_i^{Lopt} 的算法 1 以及均匀子抽样在情况(a)上的结果。每种算法我们重复循环 1000 次, 并记录所使用的 CPU 时间。由于其它三种数据集的结果相似, 因此在这省略它们以节省空间。我们不难发现, 均匀子抽样算法计算时间最短。其次, 基于 π_i^{Aopt} 的算法 1 比基于 π_i^{Lopt} 的算法需要的计算时间更长, 同时这三种算法均比不上在全数据下的 bootstrap 抽样花费的时间长, 这是符合常理的。

Table 1. $n_0 = 200$, CPU times for case (a) with n not fixed and different sampling methods
表 1. $n_0 = 200$, 情况(a)在 n 不固定和不同抽样方法下的 CPU 时间

方法	时间(秒)	n					
		100	200	300	500	700	1000
均匀		1.277	1.417	1.391	1.744	1.675	2.125
Aopt		4.501	4.717	4.969	5.328	5.645	6.022
Lopt		2.968	3.038	3.247	3.573	3.786	4.287

5.1.3. 样本量与 CPU 时间

为了进一步研究在大样本高维情况下算法 1 的性能, 我们将维数扩大至 9 维, 将参数 β 的真实值设为 $\beta_0 = (-0.5, -0.5, -0.5, -0.5, 0.5, 0.5, 0.5, 0.5, 0.5)^T$, 并考虑样本量分别为 $N = 10^4, 10^5, 10^6$ 。表 2 我们探讨了在不同抽样方法下的样本量与 CPU 时间的关系。在数据情况(a)下, 我们每次抽样方法重复循环 200 次, 同时固定第一步和第二步样本大小分别为 $n_0 = 200$ 、 $n = 1000$ 。

结果表明, 随着样本量 N 的增加, 每种抽样方法所使用的 CPU 时间都有所增加; 在相同样本量下, 最优子抽样方法相对于全数据方法大大减少了计算时间, 计算效率明显提高, 同时, 基于 π_i^{Lopt} 的算法 1 在计算上占有很大优势, 这个实验结论和表 1 的是一样的。

Table 2. $n_0 = 200, n = 1000$ and $d = 10$, case (a) CPU times under different N and different sampling methods
表 2. $n_0 = 200, n = 1000$ 和 $d = 10$, 情况(a)在 N 不同和不同抽样方法下的 CPU 时间

方法	时间(秒)	n		
		10^4	10^5	10^6
均匀		2.075	2.164	2.148
Aopt		4.732	24.461	150.817
Lopt		1.765	6.386	49.827
全数据		5.564	44.098	513.789

5.2. ϕ 未知的模拟研究

在 ϕ 未知情况下, 我们假设初始值 $\phi_0 = 2$ 。我们利用算法 2 分别研究 $\hat{\theta}$ 和 $\hat{\beta}$ 的均方误差, 进而评估算

法 2 的性能, 最后还与算法 1 进行比较。

5.2.1. 数据集分布、子样本量与 $\hat{\beta}$ 的 MSE

在本节中, 我们研究数据集分布、子样本量与 $\hat{\beta}$ 均方误差的关系, 进而评估算法 2 的性能。我们选取第一步子样本量 $n_0 = 200$, 第二步子样本大小为 $n = 500, 1000, 1500, 2000, 2500$ 。为了节省计算时间, 本节重复执行 $S = 500$ 次, 利用算法 2 计算估计值并分别求出 $\hat{\beta}$ 的均方误差 $MSE_{\hat{\beta}}$ 。为了进行比较, 我们提供了子样本大小为 $n_0 + n$ 的均匀子抽样的均方误差结果, 并使用 500 次 bootstrap 抽样计算了在全数据下估计值的均方误差。图 2 结果表明, 对于四种不同的分布, 使用最优子抽样算法比均匀子抽样方法更有优势, 并且算法 2 在基于抽样概率 π_i^{Aopt} 的 MSE 小于基于 π_i^{Lopt} 的 MSE。值得注意的是, 对于 EXP 后尾数据集, 算法 2 优于均匀抽样, 同时基于抽样概率 π_i^{Aopt} 的算法 2 与基于 π_i^{Lopt} 的性能差距不明显。

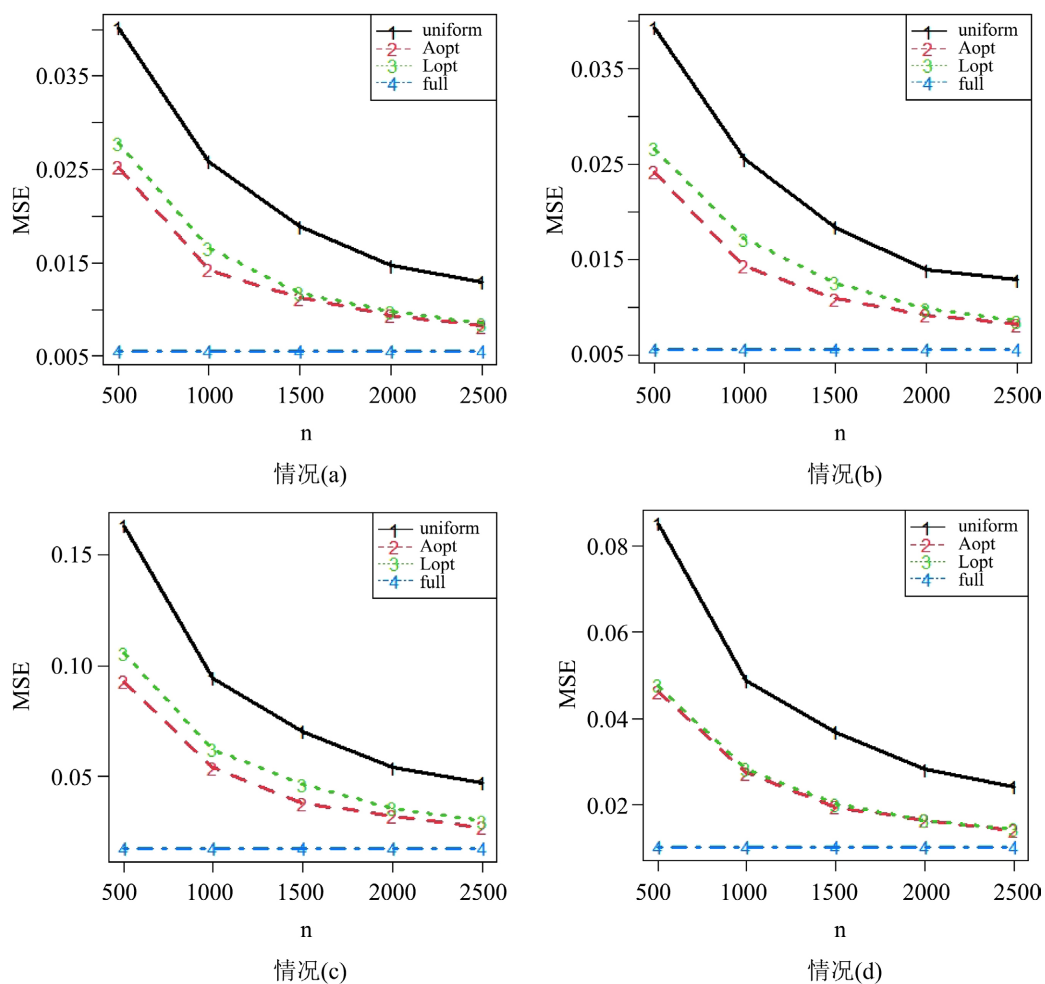


Figure 2. $n_0 = 200$, MSE of $\hat{\beta}$ for different data sets with n not fixed and different sampling methods

图 2. $n_0 = 200$, 不同数据集在 n 不固定和不同抽样方法下的 $\hat{\beta}$ MSE

5.2.2. 数据集分布、子样本量与 $\hat{\theta}$ 的 MSE

针对情况(c)和情况(d)数据集, 我们还考虑了在不同抽样方法和不同第二步子样本量下对算法 2 的影响。为了进行比较, 我们使用 $\hat{\theta}$ 的均方误差 $MSE_{\hat{\theta}}$ 作为指标。第一步子样本量、第二步子样本量、迭代次数以及抽样方法与(5.2.1)节一样。图 3 结果表明, 针对情况(c)和情况(d)的数据集, 算法 2 均优于均匀

子抽样方法。这个实验结论与前面一致。

5.2.3. $\hat{\beta}$ 与 $\tilde{\beta}$ 的 MSE 比较

我们通过比较 $\hat{\beta}$ 与 $\tilde{\beta}$ 的均方误差, 进而比较算法 1 与算法 2 的性能。在情况(a)和情况(c)下分别研究子样本量与 MSE 的关系。固定第一步子样本量 $n_0 = 200$, 第二步子样本量 $n = 300, 500, 800, 1500, 2000, 2500, 3000, 3200, 3300, 3500$ 为了缩短计算时间, 我们重复执行 $S = 500$ 次。在图 4 中, 我们发现算法 1 与算法 2 的性能差距不大, 随着子样本量的逐渐增加, 几乎没有差距。

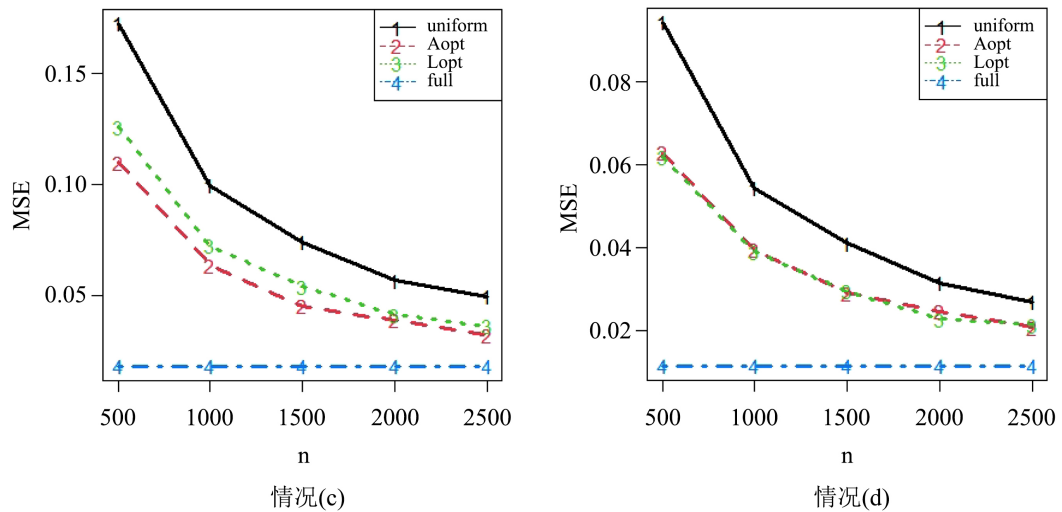


Figure 3. $n_0 = 200$, MSE of $\hat{\theta}$ for different data sets with n not fixed and different sampling methods

图 3. $n_0 = 200$, 不同数据集在 n 不固定和不同抽样方法下的 $\hat{\theta}$ MSE

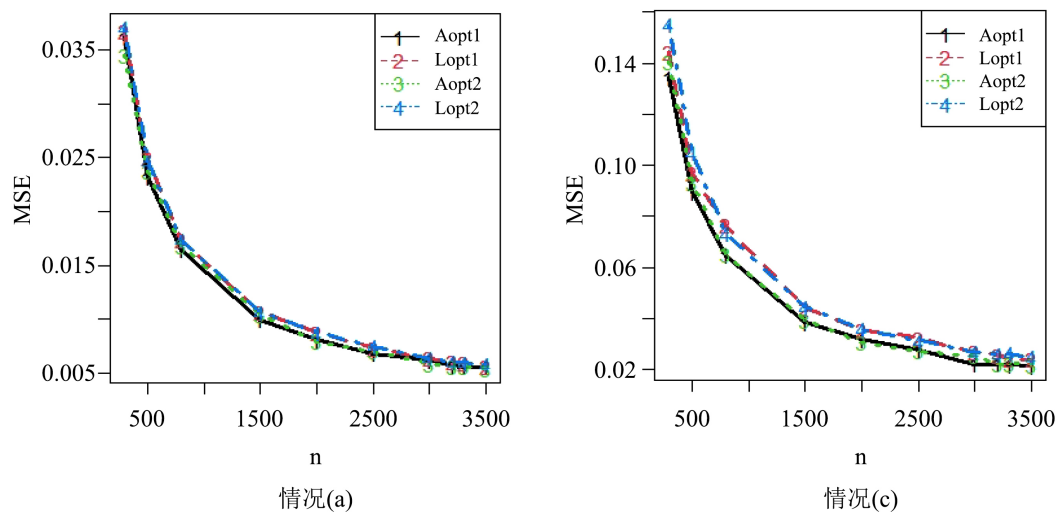


Figure 4. $n_0 = 200$, MSE for different data sets with n not fixed and different sampling probabilities

图 4. $n_0 = 200$, 不同数据集在 n 不固定和不同抽样概率下的 MSE

6. 结论

在本文中, 我们提出了两种算法分别有效地逼近 ϕ 已知和未知的 Gamma 回归在全数据下的最大似然估计。在 ϕ 已知下, 我们首先建立了一般子抽样估计量的渐近分布, 利用其渐近性质导出使渐近均方误

差最小的最优子抽样概率。根据 L -最优性准则, 我们还提出了另一种最优子抽样概率, 从而缩短了计算时间。由于最优子抽样概率取决于全数据估计量, 我们提出单参数两步算法有效地解决了此问题。在 ϕ 未知下, 我们使用在 ϕ 已知情况下所提出的两种子抽样概率, 由于子抽样概率取决于参数 ϕ 以及全数据估计量, 因此我们提出了双参数两步算法解决此问题。最后使用数值模拟分别评估这两种两步算法性能, 数值结果均表明了两种算法在大样本高维下提取有价值信息具有巨大潜力。

参考文献

- [1] Lin, N. and Xie, R. (2011) Aggregated Estimating Equation Estimation. *Statistics and Its Interface*, **4**, 73-83. <https://doi.org/10.4310/SII.2011.v4.n1.a8>
- [2] Schifano, E.D., Wu, J., Wang, C., Yan, J. and Chen, M.H. (2016) Online Updating of Statistical Inference in the Big Data Setting. *Technometrics*, **58**, 393-403. <https://doi.org/10.1080/00401706.2016.1142900>
- [3] Politis, D.N., Romano, J.P. and Wolf, M. (1999) Subsampling. Springer, New York. <https://doi.org/10.1007/978-1-4612-1554-7>
- [4] Ma, P. and Sun, X. (2015) Leveraging for Big Data Regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, **7**, 70-76. <https://doi.org/10.1002/wics.1324>
- [5] Ma, P., Mahoney, M.W. and Yu, B. (2015) A Statistical Perspective on Algorithmic Leveraging. *Journal of Machine Learning Research*, **16**, 861-911.
- [6] Ma, P., Zhang, X., Xing, X., Ma, J. and Mahoney, M. (2020) Asymptotic Analysis of Sampling Estimators for Randomized Numerical Linear Algebra Algorithms. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Vol. 108, 1026-1035.
- [7] Drineas, P., Mahoney, M. and Muthukrishnan, S. (2006) Sampling Algorithms for L_2 Regression and Applications. *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, Miami, 22-24 January 2006, 1127-1136. <https://doi.org/10.1145/1109557.1109682>
- [8] Drineas, P., Mahoney, M., Muthukrishnan, S. and Sarlos, T. (2011) Faster Least Squares Approximation. *Numerische Mathematik*, **117**, 219-249. <https://doi.org/10.1007/s00211-010-0331-6>
- [9] Dhillon, P., Lu, Y.C. and Foster, D.P. (2013) New Subsampling Algorithms for Fast Least Squares Regression. *Advances in Neural Information Processing Systems*, **1**, 360-368.
- [10] Zhu, R., Ma, P. and Mahoney, M.W. (2015) Optimal Subsampling Approaches for Large Sample Linear Regression.
- [11] Wang, H., Yang, M. and Stufken, J. (2019) Information-Based Optimal Subdata Selection for Big Data Linear Regression. *Journal of the American Statistical Association*, **114**, 393-405. <https://doi.org/10.1080/01621459.2017.1408468>
- [12] Chen, X.Y. and Xie, M.G. (2014) A Split-and-Conquer Approach for Analysis of Extraordinarily Large Data. *Statistica Sinica*, **24**, 1655-1684. <https://doi.org/10.5705/ss.2013.088>
- [13] Wang, H., Zhu, R. and Ma, P. (2018) Optimal Subsampling for Large Sample Logistic Regression. *Journal of the American Statistical Association*, **113**, 829-844. <https://doi.org/10.1080/01621459.2017.1292914>
- [14] Yao, Y. and Wang, H. (2018) Optimal Subsampling for Softmax Regression. *Statistical Papers*, **60**, 585-599. <https://doi.org/10.1007/s00362-018-01068-6>
- [15] Wang, H. and Ma, Y. (2021) Optimal Subsampling for Quantile Regression in Big Data. *Biometrika*, **108**, 99-112. <https://doi.org/10.1093/biomet/asaa043>
- [16] 李莉莉, 靳士楠, 周楷贺. 基于岭回归模型大数据最优子抽样算法研究[J]. 系统科学与数学, 2022, 42(1): 50-63.
- [17] 牛晓阳, 邹家辉. 非参数局部多项式回归估计的最优子抽样算法[J]. 系统科学与数学, 2022, 42(1): 72-84.
- [18] 王星. 大数据分析: 方法与应用[M]. 北京: 清华大学出版社, 2013.
- [19] 周围, 朱勇, 杜玉晗, 谢俊好. Gamma 分布在海面目标检测中的应用[J]. 雷达科学与技术, 2021, 19(3): 310-321.
- [20] 吴林林. 利用雨滴谱对移动双偏振雷达进行质量控制及降水估测[D]: [博士学位论文]. 南京: 南京信息工程大学, 2014.
- [21] 王俊轶. 显示画质影响要素研究[D]: [硕士学位论文]. 南京: 东南大学, 2019.
- [22] 陈超. 两参数伽马分布形状参数估计方法研究[D]: [硕士学位论文]. 淄博: 山东理工大学, 2015.
- [23] Qasim, M., Amin, M. and Amanullah, M. (2018) On the Performance of Some New Liu Parameters for the Gamma Regression Model. *Journal of Statistical Computation and Simulation*, **88**, 3065-3080. <https://doi.org/10.1080/00949655.2018.1498502>

- [24] 左卫兵, 钱莉, 谢蕾蕾. Gamma 回归模型的几乎无偏岭估计[J]. 统计与决策, 2020, 36(18): 18-21.
<https://doi.org/10.13546/j.cnki.tjyjc.2020.18.004>
- [25] Kiefer, J. (1959) Optimum Experimental Designs. *Journal of the Royal Statistical Society*, **21**, 272-319.
<https://doi.org/10.1111/j.2517-6161.1959.tb00338.x>
- [26] van der Vaart, A. (1998) Asymptotic Statistics. Cambridge University Press, London.
<https://doi.org/10.1017/CBO9780511802256>
- [27] Ferguson, T.S. (1996) A Course in Large Sample Theory. Chapman and Hall, London.
<https://doi.org/10.1007/978-1-4899-4549-5>

附录

引理 1. 如果假设 1 和 2 成立, 则

$$G_N^* - G_N = O_{P|D_N} \left(n^{-\frac{1}{2}} \right), \quad (11)$$

$$\frac{1}{N} \frac{\partial l^*(\beta)}{\partial \beta} = O_{P|D_N} \left(n^{-\frac{1}{2}} \right), \quad (12)$$

$$\text{其中 } G_N^* = \frac{1}{Nn} \sum_{i=1}^n \frac{y_i^* x_i^* (x_i^*)^T}{\phi \pi_i^* \exp \left((x_i^*)^T \hat{\beta}_{MLE} \right)}.$$

证明(引理 1): 通过计算, 可直接得出

$$E(G_N^* | D_N) = G_N \quad (13)$$

根据赫尔德不等式、假设 2 以及假设 1, 我们有

$$\begin{aligned} \text{var}(G_N^* | D_N) &= \frac{1}{n\phi^2 N^2} \sum_{i=1}^N \frac{\left[y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right]^2 \|x_i\|^4}{\pi_i} - \frac{1}{n} (G_N)^2 \\ &\leq \frac{1}{n\phi^2} \left[\frac{1}{N} \sum_{i=1}^N \left[\frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} \right]^4 \right]^{\frac{1}{2}} \left[\frac{1}{N^3} \sum_{i=1}^N \frac{\|x_i\|^8}{\pi_i^2} \right]^{\frac{1}{2}} - \frac{1}{n} (G_N)^2 \\ &= O_p \left(\frac{1}{n} \right). \end{aligned} \quad (14)$$

根据(13)式和(14)式结果, 并利用切比雪夫不等式, 可证明出(11)式。下面证明(12)式, 我们通过计算很容易得到

$$E \left[\frac{1}{N} \frac{\partial l^*(\hat{\beta}_{MLE})}{\partial \beta} \middle| D_N \right] = \frac{1}{N} \frac{\partial l(\hat{\beta}_{MLE})}{\partial \beta} = 0. \quad (15)$$

利用假设 2 和赫尔德不等式, 我们有

$$\begin{aligned} \text{var} \left\{ \frac{1}{N} \frac{\partial l^*(\hat{\beta}_{MLE})}{\partial \beta} \middle| D_N \right\} &= \frac{1}{\phi^2 N^2 n} \sum_{i=1}^N \frac{\left[y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1 \right]^2 \|x_i\|^2}{\pi_i} \\ &\leq \frac{1}{\phi^2 n} \left\{ \frac{1}{N^3} \sum_{i=1}^N \frac{\|x_i\|^4}{\pi_i^2} \right\}^{\frac{1}{2}} \cdot \left\{ \frac{1}{N} \sum_{i=1}^N \left[y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1 \right]^4 \right\}^{\frac{1}{2}} \\ &= O_p(n^{-1}). \end{aligned} \quad (16)$$

通过(15)(16)式和切比雪夫不等式, 可证出(12)式。

$$\text{证明(定理 1): 我们记 } t_i(\beta) = \frac{y_i / \exp(x_i^T \beta) + x_i^T \beta}{\phi}, \quad t_i^*(\beta) = \frac{y_i^* / \exp((x_i^*)^T \beta) + (x_i^*)^T \beta}{\phi}. \text{ 则}$$

$l^*(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{t_i^*(\beta)}{\pi_i^*}$, $l(\beta) = \sum_{i=1}^N t_i(\beta)$ 。通过计算, 我们可以直接得出

$$E\left[\frac{1}{N} l^*(\beta) \middle| D_N\right] = \frac{1}{N} l(\beta). \tag{17}$$

$$\begin{aligned} \text{var}\left[\frac{1}{N} l^*(\beta) \middle| D_N\right] &= \frac{1}{n} \left[\frac{1}{N^2} \sum_{i=1}^N \frac{t_i^2(\beta)}{\pi_i} - \left(\frac{1}{N} \sum_{i=1}^N t_i(\beta)\right)^2 \right] \\ &\leq \frac{1}{nN^2} \sum_{i=1}^N \frac{t_i^2(\beta)}{\pi_i} \\ &= O_P\left(\frac{1}{n}\right). \end{aligned} \tag{18}$$

最后一个等式是根据假设 3 得出的。根据(17)式和(18)式以及切比雪夫不等式, 在给定全数据 D_N 的条件下, 我们有 $1/Nl^*(\beta) - 1/Nl(\beta) \rightarrow 0$ 。根据定理 5.9 (Van der Vaart, 1998 [26]), 我们得到

$$\|\tilde{\beta} - \hat{\beta}_{MLE}\| = o_{P|D_N}(1). \tag{19}$$

使用泰勒定理(Ferguson, 1996 [27]), 我们有

$$0 = \frac{l_j^*(\tilde{\beta})}{N} = \frac{l_j^*(\hat{\beta}_{MLE})}{N} + \frac{1}{N} \frac{\partial l_j^*(\hat{\beta}_{MLE})}{\partial \beta^T} (\tilde{\beta} - \hat{\beta}_{MLE}) + \frac{1}{N} R_j. \tag{20}$$

其中 $l_j^*(\tilde{\beta})$ 是 $l^*(\beta)$ 关于 β_j 的偏导, 并且

$$R_j = (\tilde{\beta} - \hat{\beta}_{MLE})^T \int_0^1 \int_0^1 \frac{\partial^2 l_j^*\{\hat{\beta}_{MLE} + \mu\nu(\tilde{\beta} - \hat{\beta}_{MLE})\}}{\partial \beta \partial \beta^T} \nu d\mu d\nu (\tilde{\beta} - \hat{\beta}_{MLE}). \tag{21}$$

利用假设 3 和马尔可夫不等式, 当 $\tau \rightarrow \infty$ 时,

$$P\left(\frac{1}{\phi n N} \sum_{i=1}^n \frac{y_i^* / \exp((x_i^*)^T \beta)}{\pi_i^*} \|x_i^*\|^3 \geq \tau \middle| D_N\right) \leq \frac{1}{\phi N \tau} \sum_{i=1}^N y_i / \exp(x_i^T \beta) \|x_i\|^3 \rightarrow 0.$$

因此,

$$\begin{aligned} &\int_0^1 \int_0^1 \frac{\partial^2 l_j^*\{\hat{\beta}_{MLE} + \mu\nu(\tilde{\beta} - \hat{\beta}_{MLE})\}}{\partial \beta \partial \beta^T} \nu d\mu d\nu \\ &\leq \int_0^1 \int_0^1 \nu \left\| \frac{\partial^2 l_j^*\{\hat{\beta}_{MLE} + \mu\nu(\tilde{\beta} - \hat{\beta}_{MLE})\}}{\partial \beta \partial \beta^T} \right\| d\mu d\nu \\ &= O_{P|D_N}(N). \end{aligned} \tag{22}$$

将(19)式和(22)式带入(21)式中, 于是

$$1/NR_j = o_{P|D_N}(1). \tag{23}$$

通过(20)式、(23)式以及引理 1, 得出

$$\begin{aligned}
\tilde{\beta} - \hat{\beta}_{MLE} &= - \left[\frac{1}{N} \frac{\partial \hat{l}_j^*(\hat{\beta}_{MLE})}{\partial \beta^T} \right]^{-1} \left\{ \frac{\hat{l}_j^*(\hat{\beta}_{MLE})}{N} + o_{P|D_N}(1) \right\} \\
&= (G_N^*)^{-1} \left\{ \frac{\hat{l}_j^*(\hat{\beta}_{MLE})}{N} + o_{P|D_N}(1) \right\} \\
&= O_{P|D_N} \left(n^{-\frac{1}{2}} \right).
\end{aligned} \tag{24}$$

证明(定理 2):

$$\frac{\hat{l}_j^*(\hat{\beta}_{MLE})}{N} = \frac{1}{n} \sum_{i=1}^n \frac{[y_i / \exp((x_i^*)^T \hat{\beta}) - 1] x_i^*}{N \phi \pi_i^*} = \frac{1}{n} \sum_{i=1}^n \eta_i. \tag{25}$$

给定 D_N , $\eta_1, \eta_2, \dots, \eta_n$ 是独立同分布的。其均值

$$E(\eta_i | D_N) = 0. \tag{26}$$

通过假设 2 和赫尔德不等式, 可计算出其方差为

$$\text{var}(\eta_i | D_N) = nV_n = \frac{1}{\phi^2 N^2} \sum_{i=1}^N \frac{[y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1]^2 \|x_i\|^2}{\pi_i} = O_p(1). \tag{27}$$

同时, 对任何 $\varepsilon > 0$, 存在 $\rho > 0$,

$$\begin{aligned}
&\sum_{i=1}^n E \left\{ \left\| n^{-\frac{1}{2}} \eta_i \right\|^2 I \left(\|\eta_i\| > n^{\frac{1}{2}} \varepsilon \right) \middle| D_N \right\} \\
&\leq \frac{1}{n^{1+\frac{\rho}{2}} \varepsilon^\rho} \sum_{i=1}^n E \left\{ \|\eta_i\|^{2+\rho} I \left(\|\eta_i\| > n^{\frac{1}{2}} \varepsilon \right) \middle| D_N \right\} \\
&\leq \frac{1}{n^{1+\frac{\rho}{2}} \varepsilon^\rho} \sum_{i=1}^n E \left\{ \|\eta_i\|^{2+\rho} \middle| D_N \right\} \\
&= \frac{1}{n^2} \frac{1}{\varepsilon^\rho} \frac{1}{\phi^{2+\rho}} \frac{1}{N^{2+\rho}} \sum_{i=1}^N \frac{[y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1]^{2+\rho} \|x_i\|^{2+\rho}}{\pi_i^{1+\rho}} \\
&\leq \frac{1}{n^2} \frac{1}{\varepsilon^\rho} \frac{1}{\phi^{2+\rho}} \left\{ \frac{1}{N} \sum_{i=1}^N [y_i / \exp(x_i^T \hat{\beta}_{MLE}) - 1]^{4+2\rho} \right\}^{\frac{1}{2}} \left\{ \frac{1}{N^{3+2\rho}} \sum_{i=1}^N \frac{\|x_i\|^{4+2\rho}}{\pi_i^{2+2\rho}} \right\}^{\frac{1}{2}} \\
&\rightarrow 0.
\end{aligned}$$

最后一个等式是利用假设 2 以及赫尔德不等式得出。根据(25) (26) (27)式以及 Lindeber-Feller 中心极限定理, 在 D_N 条件下,

$$\frac{1}{N} V_n^{-\frac{1}{2}} \hat{l}_j^*(\hat{\beta}_{MLE}) = n^{-\frac{1}{2}} \left\{ \text{var}(\eta_i | D_N) \right\}^{-\frac{1}{2}} \sum_{i=1}^n \eta_i \rightarrow N(0, I).$$

从(24)式, 得出

$$\tilde{\beta} - \hat{\beta}_{MLE} = \frac{1}{N} (G_N^*)^{-1} l_j^* (\hat{\beta}_{MLE}) + o_{P|D_N}(1). \tag{28}$$

利用引理 1 结果, 我们可得到

$$\begin{aligned} (G_N^*)^{-1} - (G_N)^{-1} &= -(G_N)^{-1} (G_N^* - G_N) (G_N^*)^{-1} \\ &= O_{P|D_N} \left(n^{-\frac{1}{2}} \right). \end{aligned} \tag{29}$$

根据假设 1 和(27)式, 经验证

$$V_N = G_N^{-1} V_n G_N^{-1} = \frac{1}{n} G_N^{-1} (nV_n) G_N^{-1} = O_p(n^{-1}). \tag{30}$$

因此, 通过(28) (29) (30)式, 我们可得到

$$\begin{aligned} V_N^{-\frac{1}{2}} (\tilde{\beta} - \hat{\beta}_{MLE}) &= V_N^{-\frac{1}{2}} N^{-1} (G_N^*)^{-1} l_j^* (\hat{\beta}_{MLE}) + o_{P|D_N}(1) \\ &= V_N^{-\frac{1}{2}} G_N^{-1} N^{-1} l_j^* (\hat{\beta}_{MLE}) + V_N^{-\frac{1}{2}} \left[(G_N^*)^{-1} - G_N^{-1} \right] N^{-1} l_j^* (\hat{\beta}_{MLE}) + o_{P|D_N}(1) \\ &= V_N^{-\frac{1}{2}} G_N^{-1} V_n^{-\frac{1}{2}} V_n^{-\frac{1}{2}} N^{-1} l_j^* (\hat{\beta}_{MLE}) + O_{P|D_N} \left(n^{-\frac{1}{2}} \right). \end{aligned}$$

由于在 D_N 条件下, $N^{-1} V_n^{-\frac{1}{2}} l_j^* (\hat{\beta}_{MLE}) \rightarrow N(0, I)$ 。因此

$$V_N^{-\frac{1}{2}} (\tilde{\beta} - \hat{\beta}_{MLE}) \rightarrow N(0, I).$$

即证明出定理 2。下面利用定理 2 的结果来证明定理 3。

证明(定理 3): 利用迹的性质以及柯西 - 施瓦茨不等式可知,

$$\begin{aligned} tr(V_N) &= tr(G_N^{-1} V_n G_N^{-1}) \\ &= \frac{1}{n\phi} \sum_{i=1}^N tr \left\{ \frac{\left\{ \left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right\}^2}{\pi_i} G_N^{-1} x_i x_i^T G_N^{-1} \right\} \\ &= \frac{1}{n\phi} \sum_{i=1}^N \left\{ \frac{1}{\pi_i} \left[\left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right]^2 \|G_N^{-1} x_i\|^2 \right\} \\ &= \frac{1}{n\phi} \sum_{i=1}^N \frac{1}{\pi_i} \sum_{i=1}^N \left\{ \frac{1}{\pi_i} \left[\left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right]^2 \|G_N^{-1} x_i\|^2 \right\} \\ &\geq \frac{1}{n\phi} \sum_{i=1}^N \left\{ \left| \frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right| \cdot \|G_N^{-1} x_i\| \right\}^2. \end{aligned}$$

当且仅当 $\pi_i \propto \left| \frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right| \cdot \|G_N^{-1} x_i\|$ 等号成立, 即证明出(9)式。

$$\begin{aligned}
\text{tr}(V_n) &= \frac{1}{n\phi} \sum_{i=1}^N \text{tr} \left\{ \frac{\left\{ \left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right\}^2}{\pi_i} x_i x_i^T \right\} \\
&= \frac{1}{n\phi} \sum_{i=1}^N \left\{ \frac{1}{\pi_i} \left[\left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right]^2 \|x_i\|^2 \right\} \\
&= \frac{1}{n\phi} \sum_{i=1}^N \frac{1}{\pi_i} \sum_{i=1}^N \left\{ \frac{1}{\pi_i} \left[\left(y_i / \exp(x_i^T \hat{\beta}_{MLE}) \right) - 1 \right]^2 \|x_i\|^2 \right\} \\
&\geq \frac{1}{n\phi} \sum_{i=1}^N \left\{ \left| \frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right| \cdot \|x_i\| \right\}^2.
\end{aligned}$$

当且仅当 $\pi_i \propto \left| \frac{y_i}{\exp(x_i^T \hat{\beta}_{MLE})} - 1 \right| \cdot \|x_i\|$ 等号成立, 即证明出(10)式。