使用 Newton-Raphson 算法模拟逻辑斯蒂回归的过程

16307130308 张雨晴

一、实验原理

N 个观测的对数似然为$l(\theta)=\sum_{i=1}^{N}\log P(x_i:\theta)$。二分法中：响应变量 y=0/1，Logistic 函数

$$P(x_i:\beta)=P(Y=1\mid x_i)=\frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}$$

对数似然：

$$\begin{aligned}l(\beta)&=\log L(\beta)=\sum_{i=1}^{N}\{y_i\log P(x_i:\beta)+(1-y_i)(1-P(x_i:\beta))\}\\&=\sum\{y_i\beta^T x_i-\log(1+e^{\beta^T x_i})\end{aligned}$$

为了最大化对数似然，令微分为 0,无法求出显性表达式，我们运用 Newton-Raphson 算法。

$$\beta^{new}=\beta^{old}-\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T}\right)^{-1}\frac{\partial l(\beta)}{\partial\beta}$$

其中：

$$\begin{cases}\dfrac{\partial l(\beta)}{\partial\beta}=X^T(Y-P)\\[2ex]\dfrac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T}=-X^TWX\end{cases}$$

P 是是拟合概率的向量且第 i 个元素为 $P(x_i:\beta^{old})$；

W 是第 i 个对角元为 $P(x_i:\beta^{old})(1-P(x_i:\beta^{old}))$ 的 N*N 对角矩阵

$$\beta^{new}=\beta^{old}+(X^TWX)^{-1}X^T(Y-P)=(X^TWX)^{-1}X^TWz$$

调整后的响应变量：$z=X\beta^{old}+W^{-1}(Y-P)$

这个算法被称作 **加权迭代最小二乘（iteratively reweighted least squares）** 或者 IRLS，因为每次迭代求解加权最小二乘问题：$\beta^{new}\leftarrow\arg\min_{\beta}(z-X\beta)^TW(z-X\beta)$

二、实验步骤

（一）生成 $\hat{\beta}$

1. 生成 beta，随机生成 X，并用 logistic 函数生成 Y。

先生成一个 X 和 Y

1） $\beta=(0.5,1.2,-1.0)^T$

2） $X=(1,X_1,X_2)$，其中 $X_j\sim N(0,I_N)$：用 *rnorm(N,mean=0,sd=1)*生成

3） 由 $P(x_i:\beta)=P(Y=1|x_i)=\frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}$，生成 P 的 N*1 矩阵

4） 对于每一个 P，利用 *Y[i] <- rbinom(1,1,P[i])*生成 Y 的 N*1 矩阵

2. 对每种样本量 N，生成 R=200 组 X，用随机梯度下降法分别预测每组的 beta_hat。

1） 生成 Hessian 矩阵

P 是是拟合概率的向量且第 i 个元素为 $P(x_i:\beta^{old})$

W 是第 i 个对角元为 $P(x_i:\beta^{old})(1-P(x_i:\beta^{old}))$ 的 N*N 对角矩阵

2） 构造循环：

$$\beta^{new}=\beta^{old}+(X^TWX)^{-1}X^T(Y-P)=(X^TWX)^{-1}X^TWz$$

调整后的响应变量：$z=X\beta^{old}+W^{-1}(Y-P)$

重复地进行求解这些方程，每一次迭代时，P 改变，因此 W 和 z 也改变。

每一次循环结果的差值的长度记为 delta ，在 delta<1e-4 的时候停止循环，返回此时 beta 的值，将 beta 放入 200*3 的矩阵。

3） 每次只在 200 个样本集中随机选取用一个 X,Y。在此基础上进行迭代：改变 P,W,z 最后得到 beta。可以认为我们的方法近似于随机梯度下降方法。

（二）分析模拟实验的结果，比较样本量对模拟效果的影响。

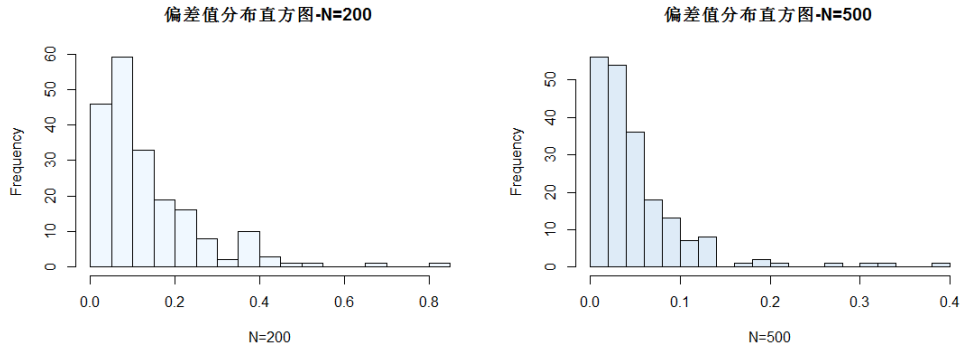
1. 构造矩阵
2. 画出偏差值的直方图
3. 画出偏差值的箱线图
4. 计算方差

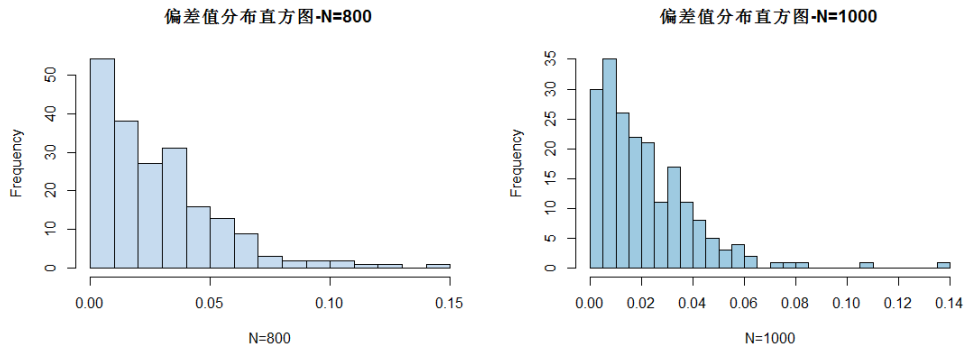
三、实验分析

1. 将每一个模拟得到的结果 $\hat{\beta}$ 减去 β=$(0.5,1.2,-1.0)^T$ 构造差值矩阵。

并算出 $\hat{\beta}-\beta$ 的长度作为总偏差值。

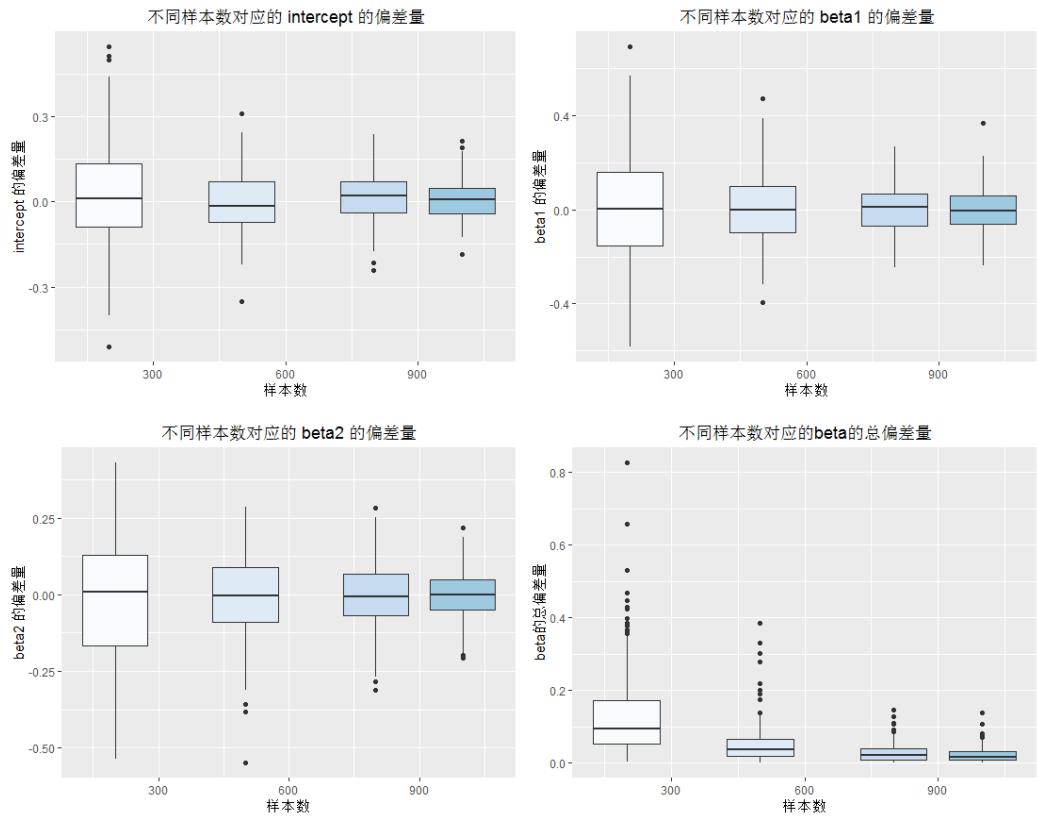
2. 分别画出样本数不同时的偏差值的直方图





从图中可以很明显的看出 N 越大，偏差值越小，且分布越集中。

3. 画出不同样本数对应的 β_j 偏差值的箱线图



图中横坐标分别为（200, 500, 800, 1000），可以看出 N 越大，箱子长度越短，说明数据分布

越集中，偏差越小，模拟效果越好。

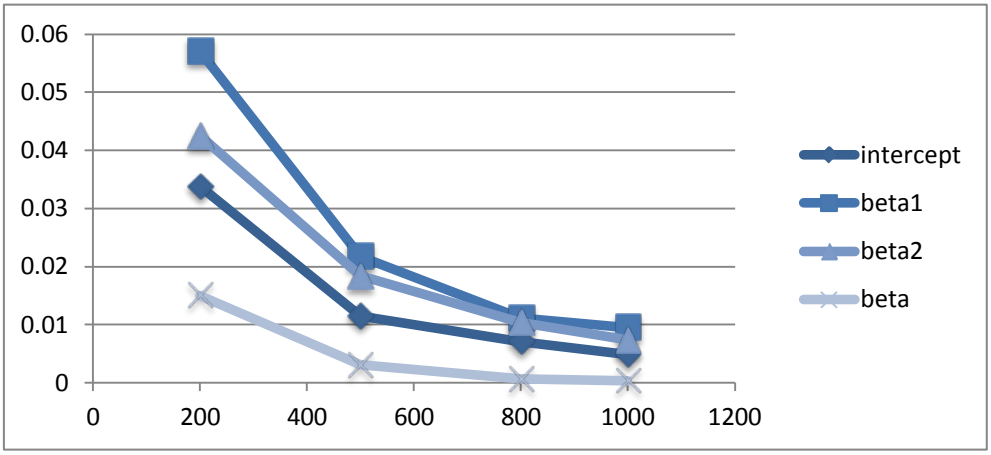
4. 为了更直观地描述结果，统计不同样本量模拟结果偏差的均值与方差如下表，并画出折线图。
   分别运用了 summary（）和 var（）的方法。

均值：

N	intercept 的偏差值平均数	Beta1 的偏差值平均数	Beta2 的偏差值平均数	Beta 的总偏差值平均数	Beta 的总偏差值中位数
200	0.01920	0.011998	0.017111	0.133621	0.093738
500	0.002576	0.0094459	0.006349	0.051524	0.035815
800	0.01744	0.004533	0.004477	0.0288440	0.0219754
1000	0.006279	0.003981	0.0038813	0.0217618	0.0167346

方差：

N	intercept 的偏差值方差	Beta1 的偏差值方差	Beta2 的偏差值方差	Beta 的总偏差值方差
200	0.03382351	0.05711232	0.04254696	0.01504221
500	0.0114778	0.02176545	0.01840237	0.003078551
800	0.007078831	0.01116282	0.0104009	0.0006454061
1000	0.004912999	0.009493325	0.007394171	0.0003706785



图表 1 不同样本数下偏差值的方差

折线图可以清晰地说明：样本数越多，偏差量的均值和方差均越小，模拟的 $\hat{\beta}$ 结果越准确。