A NEW STATISTICAL TEST FOR ZMP MODEL

杜兴兴 魏雅慧

数学与统计学院

January 24, 2021

目录

模型设定

泊松模型：

- $p(y_i|\mu) = \frac{\mu^{y_i} * e^{-\mu}}{y_i!}$, $y_i \geq 0$

泊松回归：

- $y_i|x_i \sim i.d. \; Piosson(\mu_i)$, $log(\mu_i) = x_i^T \beta$
- 当在零存在过多时，为零膨胀模型(ZIP)，由于膨胀和紧缩仅在正负号存在区别，我们在此仅讨论膨胀现象。

模型设定

零膨胀泊松(ZIP)模型：

- 假设过多零的概率为ω，则原分布的概率为1-ω。
- 混合概率分布为

$$\begin{cases} P(Y_i=0)=\omega+(1-\omega)e^{-\lambda_i} \\ P(Y_i=y_i)=(1-\omega)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad (y_i>0)! \end{cases} \tag{1.1}$$

零膨胀泊松回归：

- ω 为动态时，
  - $y_i|x_i \sim i.d.\ ZIP(\omega_i,\mu_i)$, $logit(\omega_i)=x_i^T\beta_1$, $log(\mu_i)=v_i^T\beta_2$
- ω 为常数时，
  - $y_i|x_i \sim i.d.\ ZIP(\omega_i,\mu_i)$, $\omega_i=c_i$, $log(\mu_i)=v_i^T\beta_2$

建立假设

- 检验是否为零膨胀泊松（ZIP）：

$$H_0 : \omega = 0 \quad vs \quad \omega > 0$$

统计量回顾

- Score统计量:

$$S_{score}=\frac{\left\{\sum_{i=1}^{n}\left(\frac{I_{y_i=0}}{e^{-\hat{\lambda}_i}}-1\right)\right\}}{\left\{\sum_i\left(\frac{1}{e^{-\hat{\lambda}_i}}-1\right)-\hat{\lambda}^T X[X^T diag(\hat{\lambda})X]^{-1}X^T\hat{\lambda}\right\}^{1/2}}$$

$$S_{score}\sim N(0,1)$$

- He统计量:

$$S_{new}=\frac{\left\{\sum_{i=1}^{n}\left(I_{y_i=0}-e^{-\hat{\lambda}_i}\right)\right\}}{\left\{\sum_i\left(e^{-\hat{\lambda}_i}\left(1-e^{-\hat{\lambda}_i}\right)\right)-\hat{\lambda}^T X[X^T diag(\hat{\lambda})X]^{-1}X^T\hat{\lambda}\right\}^{1/2}}$$

$$S_{score}\sim N(0,1)$$

统计量回顾

- Piosson的对数似然函数：
  - $l_1 = \sum_{i=1}^{n} \log\left(\frac{\mu^{y_i} * e^{-\mu}}{y_i!}\right)$
- ZIP的对数似然函数：
  - $l_2 = \sum_{i=1}^{n}\left[r_i \log\left(\omega + (1-\omega)\, e^{-\mu_i}\right) + (1-r_i)\log\left((1-\omega)\left(\frac{\mu^{y_i} * e^{-\mu}}{y_i!}\right)\right)\right]$
- LR Test:
  - $S_{LR} = 2[l_2(\hat{\omega}, \hat{\beta}_\mu) - l_1(0, \hat{\beta}_\mu{}')] \sim \chi_1^2$
- Wald Test:
  - $S_{Wald} = \frac{\hat{\omega}}{\hat{\sigma_\omega}^2} \sim \chi_1^2$ 或 $Z_{Wald} = \frac{\hat{\omega}}{\hat{\omega}} \sim N(0,1)$

- 检验是否为零膨胀泊松（ZIP）：

$$H_0 : \omega = 0 \quad vs \quad \omega > 0$$

- 均值变动：$y \sim Poisson(\mu)$
- 考虑四种情形：
  - 无协变量，$\mu = c$
  - 协变量服从均匀分布，$\mu = exp(\alpha - 1.4 * x)$
  - 协变量服从正态分布，$\mu = exp(\alpha - x)$
  - 多个协变量，$\mu = exp(\alpha + 0.5(x_1 + x_2))$
- 样本量：50，100，200，500
- Monte Carlo(蒙特卡罗)样本量：100

模拟介绍

- 详情见R程序
  - R Code
  - Type Error I:p-p plot
  - Results

- 均值变动：$y \sim Poisson(\mu)$
- 考虑四种情形：
  - 无协变量，$\mu = c$
  - 协变量服从均匀分布，$\mu = exp(\alpha - 1.4 * x)$
  - 协变量服从正态分布，$\mu = exp(\alpha - x)$
  - 多个协变量，$\mu = exp(\alpha + 0.5(x_1 + x_2))$
- ω变动：$y \sim ZIP(\mu, \omega)$
- 考虑三种情形：
  - $\omega = c$
  - $logit(\omega) = \alpha + x$
  - $\omega = c|\sin(4\pi x)|$
- 样本量：50，100，200，500
- Monte Carlo(蒙特卡罗)样本量：100

模拟介绍

- 详情见R程序
  - R Code
  - Type Error I:p-p plot
  - Results