第 6 讲：分类费率厘定 2 - 广义线性模型

高光远

中国人民大学 统计学院

主要内容

1. 广义线性模型
   - 指数型分布族
   - 连接函数
   - 常见的指数型分布
2. 参数估计
   - 迭代加权最小二乘法
   - 离散参数的估计
   - 极大似然估计的大样本性质
3. 预测和检验
   - 预测
   - 检验

分类费率厘定模型: 确定性 VS 随机性

- 从精算角度考虑, 费率因子需要和潜在损失相关.
- 潜在损失观测不到, 只能看到经验损失 (随机变量).
- 利用经验损失数据, 检验费率因子的统计显著性.
- 确定性模型无法进行统计检验, 需要建立随机性模型.

线性回归模型 (linear regression, linear model)

$$Y_i \overset{ind}{\sim} \mathrm{N}(\mu_i, \sigma^2) \ \text{ with } \ \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_d x_{id}, \quad i = 1, \ldots, n.$$

基本假设:

1. Y_i 互相独立.
2. 方差不随期望的改变而变化.
3. Y_i 服从正态分布.
4. Y_i 的期望 μ_i 和协变量 $\boldsymbol{x}_i$ 线性相关.

广义线性模型 (generalized linear model, GLM)

$$Y_i \overset{ind}{\sim} \text{EDF}(\mu_i, \phi) \text{ with } g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_d x_{id},\ i = 1, \ldots, n.$$

基本假设:

1. Y_i 互相独立.
2. 方差可以随期望的改变而变化.
3. Y_i 服从指数型分布族 (exponential dispersion family, EDF). 其离散系数 (dispersion)为 ϕ.
4. Y_i 的期望 μ_i 通过连接函数 (link function) g 和协变量 $\boldsymbol{x}_i$ 线性相关.

自然参数和离散参数

指数分布族的概率密度函数可以表示为

$$f(y_i;\theta_i,\phi)=\exp\left[\frac{y_i\theta_i-b(\theta_i)}{a(\phi)}+c(y_i,\phi)\right]$$

- θ_i 称为自然参数 (natural parameter), 不等于均值但和均值有关.
- ϕ 称为离散参数 (dispersion parameter), 不等于方差但和方差有关.
- 每个观察值有**不同**的自然参数, 但有**相同**的离散参数.

期望和方差

- 函数 $a(\phi)$ 通常为 ϕ/ω_i, 其中 ω_i 表示 y_i 的权重, 反映了 y_i 的相对可信度.
- 通常, 当 Y_i 表示第 i 个风险集合的**平均索赔次数**时, 权重为该风险集合的**风险单位数**; 当 Y_i 表示第 i 个风险集合的**平均索赔金额**时, 权重为该风险集合的**赔付次数**.
- Y_i 的期望为 $\mathbb{E}(Y_i) = \mu_i = b'(\theta_i)$. **期望由自然参数唯一确定**, 和 ϕ 无关.
- 函数 $c(y_i, \phi)$ 和自然参数无关, 进而和期望无关. 因此也不会影响**回归参数** β 的估计.
- Y_i 的方差为

$$\mathrm{Var}(Y_i) = a(\phi)b''(\theta_i) = \phi b''(\theta_i)/\omega_i$$

- 定义方差函数 (variance function)

$$V(\mu_i) = b''(\theta_i)/\omega_i$$

注: 连接函数的选择不影响分布.

- 期望 μ_i 通过连接函数 g, 与线性预测项 (linear predictor) η_i 关联:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_d x_{id} = \langle \beta, \boldsymbol{x}_i \rangle.$$

- 规范连接函数 (canonical link function)

$$g_c(\mu_i) = \theta_i.$$

即 g_c 为 b' 的逆函数. 当选用规范连接函数时, 自然参数等于线性预测项.

- 常见的规范连接函数有: 恒等 (正态), 对数 (泊松), 倒数 (伽马). 在 logistic 回归中, 常使用Logit 连接函数, 即

$$g(\mathbb{E}(Y_i)) = g(p_i) = \ln \frac{p_i}{1 - p_i} = \langle \beta, \boldsymbol{x}_i \rangle$$

	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
$f(y)$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$	$\frac{\mu^y\exp(-\mu)}{y!}$	$\binom{n}{y}\left(\frac{\mu}{n}\right)^y\left(1-\frac{\mu}{n}\right)^{n-y}$	$\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}y^{\nu-1}\exp\left(-\frac{\nu y}{\mu}\right)$	$\sqrt{\frac{\gamma}{2\pi y^3}}\exp\left[\frac{-\gamma(y-\mu)^2}{2\mu^2 y}\right]$
Range	$-\infty<y<\infty$	$y=0,1,2,\ldots$	$y=0,1,\ldots,n$	$y>0$	$y>0$
θ	μ	$\log(\mu)$	$\log\left(\frac{\mu}{n-\mu}\right)$	$-\frac{1}{\mu}$	$\frac{-1}{2\mu^2}$
ϕ	σ^2	1	1	$\frac{1}{\nu}$	$\frac{1}{\gamma}$
$a(\phi)$	$\phi(=\sigma^2)$	$\phi(=1)$	$\phi(=1)$	$\phi\left(=\frac{1}{\nu}\right)$	$\phi\left(=\frac{1}{\gamma}\right)$
$b(\theta)$	$\frac{\theta^2}{2}$	$\exp(\theta)$	$n\log\left(1+e^{\theta}\right)$	$-\log(-\theta)$	$-\sqrt{-2\theta}$
$c(y,\phi)$	$-\frac{1}{2}\left[\frac{y^2}{\phi}+\log(2\pi\phi)\right]$	$-\log(y!)$	$\log\binom{n}{y}$	$\nu\log(\nu y)-\log(y\Gamma(\nu))$	$-\frac{1}{2}\left[\log(2\pi y^3\phi)+\frac{1}{\phi y}\right]$
$V(\mu)$	1	μ	$\mu(1-\mu/n)$	μ^2	μ^3
$g_c(\mu)$	μ	$\log(\mu)$	$\frac{\mu}{n-\mu}$	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
$D(y,\hat{\mu})$	$(y-\hat{\mu})^2$	$2y\log\left(\frac{y}{\hat{\mu}}\right)-2(y-\hat{\mu})$	$2\left[y\log\left(\frac{y}{\hat{\mu}}\right)+(n-y)\log\left(\frac{n-y}{n-\hat{\mu}}\right)\right]$	$2\left[\frac{y-\hat{\mu}}{\hat{\mu}}-\log\left(\frac{y}{\hat{\mu}}\right)\right]$	$\frac{(y-\hat{\mu})^2}{\hat{\mu}^2 y}$

Table 2.1 *Some Exponential Family Distributions. Note that when $y=0$, $y\log(y/\hat{\mu})$ is replaced by zero (its limit as $y\to 0$).*

Version 1

- $g(\mu_i)$ 和协变量**线性相关**. 可以对 $g(\mu_i)$ 和协变量进行回归分析. 但是, μ_i 不知道, 暂用 y_i 代替.
- 利用 Taylor 展开式, 可知 $g(Y_i)$ 的方差近似为

$$\mathrm{Var}(g(Y_i)) \approx \mathrm{Var}(Y_i)g'(\mu_i)^2 = \phi V(\mu_i)g'(\mu_i)^2. \tag{1}$$

  上述方法也称为delta method.
- 数据的离散度越大, 其可信度越低. 在 $g(Y_i)$ 和协变量的回归分析中, 需要使用加权最小二乘法 (weighted least squares), 相对权重为**方差的倒数**:

$$w_i = \frac{1}{V(\mu_i)g'(\mu_i)^2}.$$

- ϕ 为常数, 在相对权重中可以省略, 不影响参数估计.

Version 1

迭代加权最小二乘法 1 (iterative re-weighted least squares, IRLS):

给定初始 $\hat{\mu}_i^0 = y_i$, 计算 $w_i^0 = 1/(V(\hat{\mu}_i^0)g'(\hat{\mu}_i^0)^2)$. 第 t 步迭代为:

1. 以 w_i^{t-1} 为权重, 通过如下加权最小二乘法求得 $\hat{\beta}^t$:

$$\arg\min_{\beta} \sum_{i=1}^{n} w_i^{t-1} \left(g(\hat{\mu}_i^{t-1}) - \langle \beta, \boldsymbol{x}_i \rangle \right)^2$$

   $\hat{\beta}^t$ 的解析解为 $\hat{\beta}^t = (X^T W X)^{-1} X^T W Y$, 其中 X 为设计矩阵 (design matrix), $W = \mathrm{diag}(w_1^{t-1}, \ldots, w_n^{t-1})$.

2. 计算新的期望 $\hat{\mu}_i^t = g^{-1}(\langle \beta^t, \boldsymbol{x}_i \rangle)$ 和新的权重:

$$w_i^t = \frac{1}{V(\hat{\mu}_i^t) g'(\hat{\mu}_i^t)^2}.$$

重复以上两步, 最终可以求出 β 的极大似然估计.

Version 2

假设有 n 个观测值, 全样本可表示为 $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i=1:n}$. 可以推出其对数似然函数为

$$l(\beta|\mathcal{D}) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(\phi, y_i) = \sum_{i=1}^{n} \frac{\omega_i \left[y_i\theta_i - b(\theta_i)\right]}{\phi} + c(\phi, y_i)$$

上式关于 β_j 求偏导

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{n}\omega_i\left(y_i\frac{\partial\theta_i}{\partial\beta_j} - b'(\theta_i)\frac{\partial\theta_i}{\partial\beta_j}\right) = \frac{1}{\phi}\sum_{i=1}^{n}\omega_i\left[y_i - b'(\theta_i)\right]\frac{\partial\theta_i}{\partial\beta_j} \tag{2}$$

其中,

$$\frac{\partial\theta_i}{\partial\beta_j} = \frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta_j} = \frac{1}{b''(\theta_i)}\frac{\partial\mu_i}{\partial\beta_j}.$$

Version 2

式 (2) 进一步等于

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{n}\frac{y_i - b'(\theta_i)}{b''(\theta_i)/\omega_i}\frac{\partial \mu_i}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{n}\frac{y_i - \mu_i}{V(\mu_i)}\frac{\partial \mu_i}{\partial \beta_j} \tag{3}$$

所以, β 的极大似然估计, 为以下方程组的解

$$\sum_{i=1}^{n}\frac{y_i - \mu_i}{V(\mu_i)}\frac{\partial \mu_i}{\partial \beta_j} = 0 \ \forall\ j \tag{4}$$

如果式 (4) 中 $V(\mu_i)$ 为已知, 且和 β 独立, 则式(4)对应于一个非线性加权最小二乘法, 其目标函数为:

$$\mathcal{S} = \sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{V(\mu_i)} \tag{5}$$

Version 2

迭代加权最小二乘法 2 (iterative re-weighted least squares, IRLS)

给定初始 $\hat{\beta}^0$, 计算 $V(\hat{\mu}_i^0) = V(g^{-1}(\langle\hat{\beta}^0, \boldsymbol{x}_i\rangle))$. 第 t 步迭代为:

1. 以 $1/V(\hat{\mu}_i^{t-1})$ 为权重, 通过非线性加权最小二乘法 (5) 求得 $\hat{\beta}^t$.
2. 计算 $V(\hat{\mu}_i^t) = V(g^{-1}(\langle\hat{\beta}^t, \boldsymbol{x}_i\rangle))$.

重复以上两步, 最终可以求出 β 的极大似然估计.

Pearson 方法

定义Pearson 统计量为

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

当样本足够大, 近似地,

$$\frac{X^2}{\phi} \sim \chi^2_{n-d-1}.$$

所以, 离散参数 ϕ 的估计为

$$\hat{\phi}_P = \frac{X^2}{n-d-1}.$$

此方法为 R 估计离散参数的方法.

偏差 (deviance)

- 定义偏差统计量 (deviance, deviance statistics)为

$$D(\hat{\beta}_{full}, \hat{\beta}) = 2\phi[l(\hat{\beta}_{full}) - l(\hat{\beta})].$$

  其中 $\hat{\beta}_{full}$ 为饱和模型 (saturated model)的参数估计.
- 在饱和模型中, 一个因变量对应一个期望参数 μ_i, 共有 n 个 (回归) 参数. 可以推出, 参数的极大似然估计为 $\hat{\mu}_i = y_i$.
- 因此, 计算 $\phi l(\hat{\beta}_{full})$ 时, 只需要知道因变量的值, 不需要知道协变量的值, $l(\hat{\beta}_{full})$ 也常记作 $l(\boldsymbol{y})$.
- 符号 $\hat{\beta}_{full}$ 主要为了和 $\hat{\beta}$ 对比, 并没有实际意义.
- 偏差统计量类似于sum of squared error (SSE), 反映了拟合优度. 可以证明, 在高斯分布条件下, Deviance=SSE.
- Deviance 和 SSE 的缺点是对过拟合不敏感.

偏差法

- 定义scaled-偏差统计量 (scaled deviance)为

$$D^*(\hat{\beta}_{full}, \hat{\beta}) = D(\hat{\beta}_{full}, \hat{\beta})/\phi = 2[l(\hat{\beta}_{full}) - l(\hat{\beta})].$$

- 关于常见指数型分布的偏差统计量, 请参考第 9页
- 可以证明 $D(\hat{\beta}_{full}, \hat{\beta})$ 和 ϕ 无关, $D^*(\hat{\beta}_{full}, \hat{\beta})$ 和 ϕ 有关. (作业)
- 若模型正确且样本足够大, $D^*(\hat{\beta}_{full}, \hat{\beta}) = D(\hat{\beta}_{full}, \hat{\beta})/\phi$ 近似服从 χ^2_{n-d-1}. 所以, 离散参数的另外一种估计方法为:

$$\hat{\phi}_D = \frac{D(\hat{\beta}_{full}, \hat{\beta})}{n-d-1}.$$

- 在线性回归中, 上式可写为:

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-d-1}$$

定义费雪信息 (Fisher information matrix)如下

$$\mathcal{I}(\beta) = \left(-\mathbb{E}\left[\frac{\partial^2 l(\beta)}{\partial \beta_l \beta_r}\right]\right)_{0 \le l, r \le d+1} \tag{6}$$

当样本足够大时, 可以证明

$$\hat{\beta} \sim \mathrm{N}(\beta, \mathcal{I}(\beta)^{-1}) \tag{7}$$

可以看到极大似然估计为无偏估计.

根据(7)可以做假设检验 (Hypothesis testing), 如

$$H_0 : \beta_r = 0.$$

在索赔频率回归模型中, 上面的假设检验即检验第 r 个分类变量和索赔频率的相关性是否统计显著.

预测

- Y_i 的期望为 μ_i. μ_i 为常数, 可通过 $\hat{\beta}, \boldsymbol{x}_i$ 进行估计

$$\hat{\mu}_i = g^{-1}(\langle \hat{\beta}, \boldsymbol{x}_i \rangle).$$

- 通过 Taylor 展开, $\hat{\mu}_i$ 的估计方差为:

$$\mathrm{Var}(\hat{\mu}_i) \approx \frac{\boldsymbol{x}_i^T \mathrm{Var}(\hat{\beta}) \boldsymbol{x}_i}{g'(\hat{\mu}_i)^2}. \tag{8}$$

- 随机变量 Y_i 也可以通过 $\hat{Y}_i = \hat{\mu}_i$ 进行预测. 其预测均方误差 (mean square error of prediction, MSEP)为

$$\mathbb{E}\left[\left(Y - \hat{Y}\right)^2\right] = \left(\mathbb{E}[Y] - \mathbb{E}[\hat{Y}]\right)^2 + \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y). \tag{9}$$

其中, 右边第一项称为偏差 (bias), 第二项称为估计方差 (estimation variance), 最后一项称为过程方差 (process variance).

Pearson residuals

定义Pearson 残差为

$$\epsilon_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}.$$

定义 scaled-Pearson 残差为

$$\epsilon_i^{SP} = \frac{y_i - \hat{\mu}_i}{\sqrt{\phi V(\hat{\mu}_i)}}.$$

可知

$$X^2 = \sum_{i=1}^{n} \left(\epsilon_i^P\right)^2$$

Deviance residuals

定义偏差残差 (deviance residuals)为

$$\epsilon_i^D = \text{sign}(y_i - \hat{\mu}_i)\sqrt{2\phi[l_i(\hat{\beta}_{full}) - l_i(\hat{\beta})]}.$$

其中, l_i 表示第 i 个样本的对数似然函数.
定义 scaled-偏差残差 (scaled-deviance residuals) 为

$$\epsilon_i^{SD} = \text{sign}(y_i - \hat{\mu}_i)\sqrt{2[l_i(\hat{\beta}_{full}) - l_i(\hat{\beta})]}.$$

可知

$$D(\hat{\beta}_{full}, \hat{\beta}) = \sum_{i=1}^{n} \left(\epsilon_i^D\right)^2$$

在理想的残差图中, 残差应该在原点两侧对称随机分布, 并没有明显的模式. 否则, 需要重新考虑**分布**的选取, 或者**连接函数**的选取.

正态检验

根据参数极大似然估计的大样本性质:

$$\hat{\beta} \sim \mathrm{N}(\beta, \mathcal{I}(\beta)^{-1})$$

可以给出每个参数的置信区间, 进而进行假设检验. 如 $H_0 : \beta_r = 0$ 或 $H_0 : \beta_r = 2$.

似然比检验

检验模型中某些协变量和因变量的相关性是否统计显著.

$H_0 : \beta_1 = \ldots = \beta_p = 0$ for $1 \leq p < d$

1. 计算完整模型的 scaled-偏差统计量 $D^*(\hat{\beta}_{full}, \hat{\beta})$
2. 计算在 H_0 下模型的 scaled-偏差统计量 $D^*(\hat{\beta}_{full}, \hat{\beta}_{H_0})$
3. 定义检验统计量

$$D^*(\hat{\beta}, \hat{\beta}_{H_0}) = D^*(\hat{\beta}_{full}, \hat{\beta}_{H_0}) - D^*(\hat{\beta}_{full}, \hat{\beta}) = 2[l(\hat{\beta}) - l(\hat{\beta}_{H_0})] \tag{10}$$

4. 在 H_0 下, $D^*(\hat{\beta}, \hat{\beta}_{H_0}) \sim \chi^2_{p_1 - p_0}$. 在 α 的显著水平下, 如果

$$D_{H_0} \geq \chi^2_{p_1 - p_0}(1 - \alpha),$$

拒绝 H_0. 这里 $p_1 = d + 1$ 为完整模型 β 参数的个数, $p_0 = d + 1 - p$ 为 H_0 模型 β 参数的个数. $p_1 - p_0 = p$.

正态检验 VS 似然比检验

- 正态检验类似于线性回归中的 t 检验, 每次只能检验一个协变量的显著性.
- 似然比检验类似于线性回归中的 F 检验, 可以同时检验多个协变量的显著性, 也可以检验 $H_0 : \beta_s = \beta_t$.

1. 阅读教材 117-125 页.
2. 完成 128 页习题 4.7.
3. 证明式 (1), 式 (8), 式 (9).
4. 证明 $D(\hat{\beta}_{full}, \hat{\beta})$ 和 ϕ 无关, $D^*(\hat{\beta}_{full}, \hat{\beta})$ 和 ϕ 有关.
5. 证明 $D(\hat{\beta}_{full}, \hat{\beta}) \geq 0$.