广义线性模型

参考教材：An Introduction to Generalized Linear Models（dobson等）

日期：2021.12

More：

GLM.

Chapter 3.

exp - family : $f(y,\theta) = s(y)t(\theta)\exp\{a(y)b(\theta)\}$

$= \exp\{a(y)b(\theta) + c(\theta) + d(y)\}$.

eg. $Poi(\theta)$: $\exp\{\log\theta \cdot y - \theta - \log y!\}$

$N(\mu,\sigma^2)$: $\exp\{\frac{\mu}{\sigma^2}y - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}\}$

$Bin(n,\pi)$: $\exp\{\log(\frac{\pi}{1-\pi})y + n\log(1-\pi) + \log\binom{n}{y}\}$.

求 $E(a(y))$, $Var(a(y))$:

$\frac{d}{d\theta}\int f(y,\theta)dy = 0 \Rightarrow \int f(y,\theta)[a(y)b'(\theta) + c'(\theta)]dy = 0$

$\Rightarrow E(a(y)) = -\frac{c'(\theta)}{b'(\theta)}$

$\frac{d^2}{d\theta^2}\int f(y,\theta)dy = 0 \Rightarrow \int f(y,\theta)[(a(y)b'(\theta)+c'(\theta))^2 + a(y)b''(\theta) + c''(\theta)]dy = 0$

$\Rightarrow E(a^2(y)) = \frac{(c'(\theta))^2 b' - c''(\theta)b'(\theta) + c'(\theta)b''(\theta)}{(b'(\theta))^3}$

$\Rightarrow Var(a(y)) = \frac{b''c' - c''b'}{(b'(\theta))^3}$.

对 exp - family, log - likelihood :

$l = a(y)b(\theta) + c(\theta) + d(y)$

$U = a(y)b'(\theta) + c'(\theta)$

① $E(U) = 0$, $Var(U) = (b')^2 Var(a(y)) = \frac{b''c' - c''b'}{b'}$.

$U' = a(y)b''(\theta) + c''(\theta)$

② $E(U') = -[\frac{c'b''}{b'} - c''] = -Var(U)$.

$J = Var(U)$.

glm: 要求: 1. 诸 Y_i 独立, 同分布族 (canonical形族)

2. g 单调可导

$$g(E(Y_i)) = x_i^T\beta \qquad \beta_{p\times1},\ Y_{N\times1},\ \text{通常 } p<N$$

令 $\mu_i = E(Y_i)$

Chapter 4.

优化技术: $U(\theta)=0$, 牛顿: $\theta^{(m)} = \theta^{(m-1)} - \dfrac{U(\theta^{(m-1)})}{U'(\theta^{(m-1)})}$

Method of Scoring: $\theta^{(m)} = \theta^{(m-1)} + \dfrac{U(\theta^{(m-1)})}{J(\theta^{(m-1)})}$

Standard error of $\hat\theta$: $se(\hat\theta) = \sqrt{\dfrac{1}{J}}$

~~参数估计推导:~~

$$\cancel{\mu_i = E(Y_i),\quad g(\mu_i) = X_i^T\beta,\quad \eta_i = X_i^T\beta}$$

$$\cancel{l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)}$$

$$\cancel{① \quad l = \sum_{i=1}^N l_i = \sum_{i=1}^N [y_i b(\theta_i) + c(\theta_i) + d(y_i)]}$$

$$\cancel{U = \sum_{i=1}^N U_i = \sum_{i=1}^N [y_i b'(\theta_i) + c'(\theta_i)]}$$

估计推导: $Y_i \sim f(\theta_i) = \exp\{\cancel{a(y_i)}\,y_i b(\theta_i) + c(\theta_i) + d(y_i)\}$

$$\mu_i = E(Y_i) = -\frac{c'(\theta_i)}{b'(\theta_i)}$$

$$\eta_i = X_i^T\beta$$

$$g(\mu_i) = \eta_i$$

$\theta_i \overset{E}{\leftrightarrow} \mu_i \overset{g}{\leftrightarrow} \eta_i \leftrightarrow \beta \rightarrow$ 最终参数.

① $l_i = \cancel{a(y_i)}\,y_i b(\theta_i) + c(\theta_i) + d(y_i)$

求 $U_{p\times1}$: $U_j = \dfrac{\partial l}{\partial \beta_j} = \sum_{i=1}^N \dfrac{\partial l_i}{\partial \beta_j}$

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = ABCD.$$

$$A = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i)$$

$$B = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = +\frac{(b')^2}{b''c' - c''b'} = \frac{1}{b'(\theta_i)\, Var(Y_i)}.$$

$$C = \frac{1}{\frac{\partial \eta_i}{\partial \mu_i}} \quad 与 g 有关$$

$$D = x_{ij}$$

$$\Rightarrow U_j = \sum_{i=1}^{N}\left[\frac{y_i - \mu_i}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right], \quad U = (U_1, \cdots U_p)$$

② 求 J

$$J_{jk} = E(U_j U_k) = E\left[\sum_{i=1}^{N}\left(\frac{y_i - \mu_i}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i} X_{ij}\right)\sum_{l=1}^{N}\left(\frac{y_l - \mu_l}{Var(Y_l)}\frac{\partial \mu_l}{\partial \eta_l} X_{lk}\right)\right]$$

$$= E\left[\sum_{i=1}^{N}\frac{(y_i - \mu_i)^2}{(Var(Y_i))^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 X_{ij}X_{ik}\right]$$

注：$\mu_i = E(Y_i)$ 又 Y_i 独立，故交叉项全 0.

$$= \sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{Var(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

$$\Rightarrow J = X^T W X; \quad W = diag\left(\frac{1}{Var(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right)_{N\times N}, \quad X : N\times p.$$

③ 记 $b = \hat{\beta}$

$$b^{(m)} = b^{(m-1)} + \cancel{\frac{U(b^{(m-1)})}{J}}\ J^{-1}(b^{(m-1)})\, U(b^{(m-1)})$$

④ Iterative weighted least squares:

$$\underbrace{J(b^{(m-1)})}_{X^TWX} b^{(m)} = \underbrace{J(b^{(m-1)})b^{(m-1)} + U(b^{(m-1)})}_{X^TWZ_{N\times 1}}.$$

$$Z_{N\times 1}, \quad z_i = \underbrace{\sum_{k=1}^{p} x_{ik} b_k^{(m-1)}}_{X_i^T b^{(m-1)}} + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}.$$

$$\Rightarrow X^TWXb^{(m)} = X^TWZ$$

Chapter 5.

思想：$S_{p\times 1}$，则 $(S-E(S))^T V^{-1}(S-E(S)) \sim \chi^2(p)$.

例：对 Score function：$U_j = \sum_{i=1}^{N}\left[\frac{Y_i-\mu_i}{Var(Y_i)} x_{ij}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\right]$

$U=(U_1, \cdots U_p)$.

① 有 $U^T J^{-1} U \to \chi^2(p)$. [Score]

② 由Taylor：$U(\beta) \approx \underset{0}{U(b)} - J(b)(\beta-b)$

$\Rightarrow b-\beta = J^{-1}U$ (渐近)

$\Rightarrow E[(b-\beta)(b-\beta)^T] = E(J^{-1}UU^TJ^{-1}) = J^{-1}$.

又 $E(b-\beta)=0$ (渐近)

$\Rightarrow (b-\beta)^T J(b-\beta) \to \chi^2(p)$. [Wald]

一元时有 $b \sim N(\beta, J^{-1})$

③ 令 β_{max} 为饱和模型参数，$b_{max}=\hat{\beta}_{max}$，有m个参数

由Taylor：$l(\beta)-l(b) = -\frac{1}{2}(\beta-b)^T J(b)(\beta-b)$.

$\Rightarrow 2(l(b)-l(\beta)) \to \chi^2(p)$

定义 deviance：$D = 2[l(b_{max})-l(b)]$

$= \underset{\chi^2(m)}{\underbrace{2(l(b_{max})-l(\beta_{max}))}} - \underset{\chi^2(p)}{\underbrace{2(l(b)-l(\beta))}} + \underset{V\to 0}{\underbrace{2(l(\beta_{max})-l(\beta))}}$

$\to \chi^2(m-p, V)$

注：Y_i 正态 $\Rightarrow D\sim\chi^2$ exactly

下面推 N, Poi, Bin 的 Dev.

1. $N(\mu, \sigma^2)$

模型：$E(Y_i)=\mu_i = x_i^T\beta$, $Y_i \sim N(\mu_i, \sigma^2)$

饱和模型：μ_i 都可以不同 $\Rightarrow \hat{\mu_i} = y_i$

$$l = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\mu_i)^2 - \frac{1}{2}N\log(2\pi\sigma^2)$$

得 $l(b_{max}) = -\frac{1}{2}N\log(2\pi\sigma^2)$

对于到底模型， $b = (X^TX)^{-1}X^TY$

$$l(b) = -\frac{1}{2\sigma^2}\Sigma(y_i - x_i^Tb)^2 - \frac{1}{2}N\log(2\pi\sigma^2)$$

$$\Rightarrow D = \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \hat{\mu_i})^2.$$

$$= \frac{1}{\sigma^2}(y-Xb)^T(y-Xb) = \frac{1}{\sigma^2}y^T(I-P_X)y \sim \chi^2(N-p).$$

可以用 $\frac{1}{N-p}y^T(I-P_X)y$ 估 σ^2.

2. $Poi(\lambda_i)$

饱和：λ_i 可以变 $\Rightarrow \hat{\lambda_i} = y_i$

$$l = \sum_{i=1}^{N}[y_i\log\lambda_i - \lambda_i - \log y_i!]$$

$$\Rightarrow l(b_{max}) = \sum_{i=1}^{N}[y_i\log y_i - y_i - \log y_i!]$$

模型：用 b 算 $\hat{\lambda_i}$，故 $\hat{y_i} = \hat{\lambda_i}$

$$\Rightarrow D = 2\sum_{i=1}^{N}\left[y_i\log\frac{y_i}{\hat{y_i}} - (y_i - \hat{y_i})\right]$$，可与 $\chi^2(N-p)$ 比较.

3. $Bin(n_i, \pi_i)$

$$l = \sum_{i=1}^{N}\left[\log\pi_i \cdot y_i - y_i\log(1-\pi_i) + n_i\log(1-\pi_i) + \log\binom{n_i}{y_i}\right]$$

饱和：$\hat{\pi_i} = \frac{y_i}{n_i}$，又对到模，$\hat{y_i} = n_i\hat{\pi_i}$.

$$\Rightarrow D = 2\sum_{i=1}^{N}\left[y_i\log\left(\frac{y_i}{\hat{y_i}}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y_i}}\right)\right].$$

Hypothesis testing

除了 $(\hat{\beta}-\beta)^T J(\hat{\beta}-\beta) \to \chi^2(p)$

$U^T J^{-1} U \to \chi^2(p)$ 外

还可以用嵌套检验.

$$H_0: \beta = \beta_0 = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} \quad H_1: \beta = \beta_1 = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad q < p < N$$

用 Dev 偏差: $\Delta D = D_0 - D_1 = 2(l(b_1) - l(b_0)) \sim \chi^2(p-q)$

eg. 正态 (处理 σ^2)

① 算 D_0, D_1, $D_0 = \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \hat{\mu}_i(0))^2$

$D_1 = \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \hat{\mu}_i(1))^2$

M_1 拟合好: $D_1 \sim \chi^2(N-p)$

M_0 拟合也好: $D_0 \sim \chi^2(N-q)$

$\Rightarrow \Delta D \sim \chi^2(p-q)$

令 $F = \frac{D_0 - D_1}{p-q} \Big/ \frac{D_1}{N-p} \sim F(p-q, N-p)$ (H_0下).

Chapter 7. 二元模型

① $Z_i \sim Ber(\pi_i)$

$p(Z_1, \cdots Z_n) = \prod_{i=1}^{n} \pi_i^{Z_i}(1-\pi_i)^{1-Z_i} = \exp\left\{\sum_{i=1}^{n}\left(Z_i \log\frac{\pi_i}{1-\pi_i} + \log(1-\pi_i)\right)\right\}$ 可以 glm.

② $Y_i \sim Bin(n_i, \pi_i)$

$l(\pi_1, \cdots \pi_N) = \sum_{i=1}^{N}\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) + \log\binom{n_i}{y_i}\right]$

Y_i 常表示 n_i 次中 Success 次数

扫描全能王 创建

令 $P_i = \frac{Y_i}{n_i}$，则 $E(P_i) = \pi_i$，建模：$g(\pi_i) = x_i^T\beta$.

g 的选取：(1) $\Phi^{-1}(\cdot)$

(2) $\log\left(\frac{\pi_i}{1-\pi_i}\right)$

(3) $\log\left(-\log(1-\pi_i)\right)$.

例如取 (2)，则 $l = \sum_{i=1}^{N}\left[y_i(\beta_1+\beta_2 x_i) - n_i\log\left[1+\exp(\beta_1+\beta_2 x_i)\right] + \log\binom{n_i}{y_i}\right]$ （$X_{N\times 2}$）

max l 得 $\hat\beta_1, \hat\beta_2$ ⇒ 得 $\hat\pi_i$ ⇒ $\hat y_i = n_i\hat\pi_i$ ⇒ $D \sim \chi^2(N-2)$.

除了 mle，还可以 min $S_w = \sum_{i=1}^{N}\frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1-\pi_i)}$

得到 $X^2 = \sum_{i=1}^{N}\frac{(y_i-n_i\hat\pi_i)^2}{n_i\hat\pi_i(1-\hat\pi_i)}$，与 D 渐近.

证：$s\log\frac{s}{t} = (s-t) + \frac{1}{2}\frac{(s-t)^2}{t} + \cdots$

$$\Rightarrow D = 2\sum_{i=1}^{N}\left[y_i\log\frac{y_i}{n_i\hat\pi_i} + (n_i-y_i)\log\frac{n_i-y_i}{n_i-n_i\hat\pi_i}\right]$$

= 展到二阶 = X^2.

⇒ $X^2 \sim \chi^2(N-p)$ 且好于 D.

但当 covarite pattern 中的样本量太少时不好.

定义（最小模型）：所有系数都一样.

eg. $Bin(n_i, \pi_i)$ 中，$\tilde\pi_i = \Sigma y_i / \Sigma n_i$.

令 $C = 2(l(\hat\pi) - l(\tilde\pi)) \to \chi^2(p-1)$

pesudo $R^2 = \frac{l(\tilde\pi) - l(\hat\pi)}{l(\tilde\pi)}$.

$AIC = -2l(\hat\pi) + 2p$.

residuals： 设有 m 个 covariate pattern，Y_k，n_k，$\hat{\pi}_k$，$k=1,\cdots m$

pearson 残差： $X_k = \frac{y_k - n_k\hat{\pi}_k}{\sqrt{n_k\hat{\pi}_k(1-\hat{\pi}_k)}}$．，由此：$X^2 = \sum_1^m X_k^2$．

标准化残差： $r_{pk} = \frac{X_k}{\sqrt{1-h_k}}$，$h_k$ 为 $(P_X)_{kk}$．

Dev 残差： $d_k = \text{sign}(y_k - n_k\hat{\pi}_k)\sqrt{2\left[y_k \log \frac{y_k}{n_k\hat{\pi}_k} + (n_k - y_k)\log \frac{n_k - y_k}{n_k - n_k\hat{\pi}_k}\right]}$

由此 $D = \sum_1^m d_k^2$.

$r_{Dk} = \frac{d_k}{\sqrt{1-h_k}}$.

$r_{pk}, r_{Dk} \to N(0,1)$.

Chapter 9.

$Poi(\mu)$： $f(y, \mu) = \frac{\mu^y e^{-\mu}}{y!}$，$y=0,1,\cdots$

Poisson Regression：

$Y_i \sim Poi(\mu_i)$，$\mu_i = E(Y_i)$，$\eta_i = x_i^T \beta$

~~$\log(\mu_i) = \eta_i$~~ $\theta_i = \frac{\mu_i}{n_i}$

$\log(\theta_i) = \eta_i$

等价地： $Y_i \sim Poi(n_i e^{x_i^T\beta})$.

natural link function： $\log \mu_i = \underbrace{\log n_i}_{\text{offset}} + x_i^T\beta$.

系数 interpret： x_{ij} 增加 1 后 条件均值 的比值 为 e^{β_j}

$RR = \frac{E(Y_i \mid +1)}{E(Y_i \mid +0)} = e^{\beta_j}$.

fitted value: $\hat{Y}_i = \hat{\mu}_i = n_i e^{x_i^T b}$.

Pearson 残差: $r_i = \dfrac{o_i - e_i}{\sqrt{e_i}}$.

$$r_{pi} = \frac{o_i - e_i}{\sqrt{e_i}\sqrt{1-h_i}}, \quad \chi^2 = \sum r_i^2.$$

$$D = 2\sum\left[o_i \log\frac{o_i}{e_i} - (o_i - e_i)\right].$$

$$d_i = \mathrm{sign}(o_i - e_i)\sqrt{2\left[o_i \log\frac{o_i}{e_i} - (o_i - e_i)\right]}, \quad D = \sum d_i^2.$$

列联表检验（以三维为例）：

① 自由生长： $f(y) = \prod_i \prod_j \prod_k \dfrac{(n\theta_{ijk})^{y_{ijk}}}{y_{ijk}!} e^{-n\theta_{ijk}}$

② 给定总量： $f(y|n) = n! \prod_i \prod_j \prod_k \dfrac{\theta_{ijk}^{y_{ijk}}}{y_{ijk}!}$.

③ 给定每个边际： $f(y \mid \cancel{y_{\cdot\cdot 1}, \dots, y_{K\cdot\cdot}}\ y_{1\cdot\cdot}, y_{2\cdot\cdot}, \dots, y_{I\cdot\cdot}) = \prod_i y_{i\cdot\cdot} \prod_j \prod_k \dfrac{\theta_{ijk}^{y_{ijk}}}{y_{ijk}!}$.

Log-linear Model: $\log E(Y_i) = c + X_i^T \beta$.

求解：看成 Poisson Regression 然后验 ~~goodness of fitness~~ goodness of fit 即可。

eg. A, B, C, Freq.

(1) H_0: A, B, C 独立， Freq ~ A+B+C

(2) H_0: A, B 独立， Freq ~ AC+BC.

(3) H_0: A, B, C 都相关 但 相关程度不受第三方影响： Freq ~ AB+BC+CA.

等价地，可以将 C 作为 response，验 goodness of fit (logistic)

(1) C ~ 1 $\Leftrightarrow$ Freq ~ A+B+C+AB.

(2) C ~ A （C, B 独立） $\Leftrightarrow$ Freq ~ AB+AC

(3) C ~ A+B $\Leftrightarrow$ Freq ~ AB+BC+CA

(4) C ~ AB $\Leftrightarrow$ Freq ~ ABC.

Chapter 8

Nominal logistic regression

π_j, $j=1,\cdots,J$

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = x_j^T\beta_j + \beta_{0j}$$

$$\Rightarrow \hat{\pi}_1 = \frac{1}{1+\sum_{j=2}^{J} e^{x_j^T\hat{\beta}_j}}, \quad \hat{\pi}_j = \frac{e^{x_j^T\hat{\beta}_j}}{1+\sum_{j=2}^{J} e^{x_j^T\hat{\beta}_j}}.$$

Pearson χ^2 残差： $r_i = \frac{o_i - e_i}{\sqrt{e_i}}$. $\quad$, $X^2 = \sum_{i=1}^{N} r_i^2$.

$Dev = 2(l(b_{max}) - l(b)) \longrightarrow \chi^2(N-p)$

伪 $R^2 = \frac{l(b_{min}) - l(b)}{l(b_{min})}$.

$AIC = -2l(b) + 2p$.

$C = 2(l(b) - l(b_{min})) \longrightarrow \chi^2(p-(J-1))$ 对每个方程除一个截距.

参数解释： $\cancel{令\ OR_j = \frac{\pi_j}{1-\pi_j} \Big/ \frac{\pi_1}{1-\pi_1}}$

$\quad$ 则 $\cancel{\log(OR_j) = e^{\beta_j}\ (x增1)}$.

Ordinal Logistic Regression.

1. proportional odds model.

$$\log\frac{\pi_1+\cdots+\pi_j}{\pi_{j+1}+\cdots+\pi_J} = \beta_{0j} + x_j^T\beta \quad \text{(对不同的 j，斜率都相同)}.$$

性质：1. 合并类后，β 不变 (collapsibility)

$\quad$ 2. 把类反一下，仅 β 的符号变.

2. Adjacent categories logit model.

$$\log\frac{\pi_j}{\pi_{j+1}} = x_j^T\beta_j$$

可简化为 $\log\frac{\pi_j}{\pi_{j+1}} = x_j^T\beta$ （斜率与 j 无关）.

3. Continuation ratio logit model.

用 $\frac{\pi_1+\cdots+\pi_{j-1}}{\pi_j}$ 或 $\frac{\pi_j}{\pi_{j+1}+\cdots+\pi_J}$ 来 model.