M5 - 2018.10.11

GLM - common core

GLM: 3 ingredients

1) $Y_i \sim f(y_i; \theta_i, \phi, w_i)$ — of interest: θ_i

$$= \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i) \right)$$

ϕ ← nuisance

$\mu_i = E(Y_i) = b'(\theta_i)$

$\sigma_i^2 = Var(Y_i) = \underbrace{b''(\theta_i)}_{V(\mu_i)} \cdot \frac{\phi}{w_i}$

$V(\mu_i)$ ← variance function

2) $\eta_i = x_i^T \beta$, $\underset{n \times p}{X} = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$ ← full rank (p)

3) $\underset{\text{link}}{\eta_i = g(\mu_i)} \iff \underset{\text{response}}{\mu_i = h(\eta_i)}$
- require one-to-one
- twice differentiable

- N: $\eta_i = \mu_i$
- bin: $\mu_i = p_i\ (n_i = 1)$: $\eta_i = \ln\left(\frac{p_i}{1-p_i}\right)$
- Poisson: $\mu_i = \lambda_i$: $\eta_i = \ln \lambda_i$

gamma: $\eta_i = \ln \mu_i$, $-\frac{1}{\mu_i}$ (canonical link)

Likelihood inference

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

$$\mu_i = b'(\theta_i) \qquad \eta_i = x_i^T \beta$$

$$\begin{tikzcd}
\theta_i \arrow[r] & \mu_i \arrow[r, leftrightarrow] & \eta_i \arrow[r] \arrow[ll, bend left, "\theta_i = \eta_i"] & \beta
\end{tikzcd}$$

θ_i: canonical parameter

μ_i: mean of y_i, $E(Y_i)$

η_i: linear predictor

β: regression parameter

for canonical link

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$$

Loglikelihood

canonical link

$$\sum_{i=1}^{n} y_i \underset{\eta_i}{\theta_i} = \sum_{i=1}^{n} y_i \underbrace{x_i^T \beta}_{\sum \beta_j x_{ij}} = \sum_j \beta_j \sum_i x_{ij} y_i$$

$$l(\beta) = \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} \left(\frac{w_i}{\phi} \left(y_i \theta_i - b(\theta_i) \right) + c(y_i, \phi, w_i) \right)$$

Score function

$$s(\beta) = \frac{\partial l}{\partial \beta}$$

$$s(\beta) = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta} = \sum_{i=1}^{n} s_i(\beta)$$

$$\underbrace{\left(\frac{w_i}{\phi}\left(y_i\theta_i - b(\theta_i)\right) + c(y_i, \phi, w_i)\right)}$$

$\mu_i = b'(\theta_i)$ $\qquad$ $\eta_i = x_i^T\beta$

$$\frac{\partial l_i}{\partial \beta} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta}$$

$$\frac{w_i}{\phi}\left(y_i - \underbrace{b'(\theta_i)}_{\mu_i}\right)$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = Var(Y_i) \cdot \frac{w_i}{\phi}$$

depends on the link function $h'(\eta_i)$

$$\frac{\partial x_i^T\beta}{\partial \beta} = x_i$$

$$\frac{\partial l_i}{\partial \beta} = (y_i - \mu_i) \cdot \frac{w_i}{\phi} \cdot \frac{\phi}{w_i} \cdot \frac{1}{Var(Y_i)} \, h'(\eta_i) \, x_i$$

$$S(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i) \cdot x_i \, h'(\eta_i)}{Var(Y_i)}$$

$$S(\beta) = X^T D \Sigma^{-1} (y - \mu)$$

$D = diag(h'(\eta_i))$ $\qquad$ $\Sigma^{-1} = [diag(Var(Y_i))]^{-1}$

canonical link
$h'(\eta_i) = b''(\theta_i) = Var(Y_i)\frac{w_i}{\phi}$

$$S(\beta) = \sum (y_i - \mu_i) x_i \frac{w_i}{\phi}$$

when $w_i = \phi = 1$ as for Poisson and Bernoulli, then likelihood eq. are

suff. stat. $\rightarrow \sum x_{ij} y_i = \sum x_{ij} \mu_i$

$\underbrace{\text{the likelihood eq. equate the sufficient statistics for the model param. to their expected value}}$

added!

Observe: $s(\beta)$ depends on the distr. of Y_i only through μ_i and $Var(Y_i)$.

Expected Fisher information matrix $F(\beta)$

$$F(\beta) = Cov(s(\beta)) \underset{\uparrow}{=} \sum_{i=1}^{n} \underbrace{Cov(s_i(\beta))}_{F_i(\beta)} \quad \text{and} \quad Cov(s_i(\beta)) = E(s_i(\beta) s_i(\beta)^T)$$

$p \times p$ indep. $i = 1, .., n$ obs   since $E(s_i(\beta)) = 0$

because: $$E(s_i(\beta)) = E\left(\frac{(Y_i - \mu_i) x_i \cdot h'(\eta_i)}{Var(Y_i)}\right) = \frac{x h'(\eta_i)}{Var(Y_i)} [\underbrace{E(Y_i)}_{\mu_i} - \mu_i) = 0$$

We will look at element (h, l) of $F_i(\beta)$, and use

$$\frac{\partial l_i}{\partial \beta_h} = \frac{(y_i - \mu_i) \cdot x_{ih}}{Var(Y_i)} h'(\eta_i)$$

$$F_i(\beta)_{[h,l]} = E\left(\frac{\partial l_i}{\partial \beta_h} \cdot \frac{\partial l_i}{\partial \beta_l}\right)$$

$$= E\left[\frac{(Y_i - \mu_i) \cdot x_{ih} \cdot h'(\eta_i)}{Var(Y_i)} \cdot \frac{(Y_i - \mu) \cdot x_{il} \cdot h'(\eta_i)}{Var(Y_i)}\right]$$

$$= \frac{h'(\eta_i)^2 \cdot x_{ih} \cdot x_{il}}{Var(Y_i)^2} \underbrace{E[(Y_i - \mu_i)^2]}_{Var(Y_i)} = \frac{x_{ih} \cdot x_{il} \cdot [h'(\eta_i)]^2}{Var(Y_i)}$$

$$F(\beta)_{[h,l]} = \sum_{i=1}^{n} \frac{x_{ih} \cdot x_{il} [h'(\eta_i)]^2}{Var(Y_i)}$$

$$F(\beta) = X^T W X \qquad \text{wher } W = diag\left(\frac{h'(\eta_i)^2}{Var(Y_i)}\right)$$

Alt: $\left.\begin{array}{l} D = diag(h'(\eta_i)) \\ \Sigma = diag(Var(Y_i)) \end{array}\right\} \quad D \Sigma^{-1} D = W$

When canonical link: $H(\beta) = -\frac{\partial^2 \ell}{\partial\beta\partial\beta^T}$

$$\underbrace{E(H(\beta))}_{F(\beta)} = H(\beta)$$

PARAMETER ESTIMATION:

→ can be done seperately for β and ϕ

(see notes on orthogonal parameters – advanced topic)

$\hat{\beta}$ is found by solving $S(\hat{\beta}) = 0$ using the Fisher scoring ← $\underbrace{\text{need formules for } S(\beta) \text{ and } F(\beta)}$

↓

we get Iterative Reweighted Least Squares

IRWLS

method