#5 - F2017

Estimation for GLMs

Summary

- Exponential family
- Components of GLM
- IWLS

Reading

- DB Chapter 4
- MN Chapter 2

Review: Exponential Family

- $Y_i \sim$ Exponential family
  - Density;

$$f_i(y_i|\theta_i, \phi \ = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi)\right\},$$

where $\mu_i \equiv E[Y_i] = b'(\theta_i)$ and $V(Y_i) = b''(\theta_i)a_i(\phi)$.

  - Link between μ_i and η_i (linear term)

$$g(\mu_i) = \eta_i = \mathbf{X}_i\beta = \sum_{j=1}^{p} X_{ij}\beta_j$$

(Very often $X_{i1} = 1$ as intercept).

- Examples of CLMs (with canonical link, i.e. $\theta = \eta$) → Variance function not variance

Model	μ	$\eta = g(\mu)$	$V(\mu)$	$a(\phi$
Linear	mean	μ	1	σ^2
Logistic	Prob. of Success p	$\log \frac{\mu}{1-\mu}$	$\mu(1-\mu)$	$\frac{1}{m}$
Poisson	Expected Count $\lambda = \mu$	$\log(\mu)$	μ	1

← Consider Y/m as the response, $Y \sim Bin(m, p)$

For the Binomial family the distribution of y_i/m_i is used.

*HW Find GLM's implemented in R and SAS

Estimation for GLMs

- Maximum Likelihood Estimation (MLE) for β.
  - Log-likelihood

$$\log L = \sum_{i=1}^{n} \left\{ \frac{Y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c_i(Y_i, \phi \right\} = \sum_{i=1}^{p} \log L_i,$$

  - Score equation

$$U_j(\beta) = \frac{\partial}{\partial \beta_j} \log L = 0, \text{ for } j = 1, \quad , p.$$

  - There is no β_j in the form of $\log L$. Then, how to estimate β_j?
    - * Chain rule !

$$\theta_i \longrightarrow \mu_i \longrightarrow \eta_i \longrightarrow \beta$$

In the Canonical form, $l(\beta) = \log L$ is a function of θ_i orignally, Since $\theta_i = g(\mu_i)$, $\mu_i = g^{-1}(\eta_i) = h(\eta_i)$, $\eta_i = x_i^T\beta$, then $l(\beta)$ is a function of β

A Question Since $\theta_i = \eta_i = x_i^T\beta$, why not use $\theta_i \longrightarrow \eta_i \longrightarrow \beta$?

Answer: using $\mu_i = E(Y_i)$, it is convenient to formulate IWLS late

Then, the score functions can be written as:

$$U_j(\beta) = \frac{\partial}{\partial \beta_j} \log L = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \log L_i$$
$$= \sum_i \frac{1}{V_i (\partial \eta_i / \partial \mu_i)^2} (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij}$$

see proofs below

Estimation (continued)

Proof: For ith obs, let

$l_i = l_i(Y_i; \theta_i; \phi) = \operatorname{Log} L_i = \frac{[Y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i; \phi)$

where $\theta_i = g(\mu_i) = \eta_i$,

$\mu_i = E(Y_i) = h(\theta_i) = h(\eta_i) = g^{-1}(\eta_i)$ (i.e. $h = g^{-1}$),

$\eta_i = x_i^T \beta$.

Recall $\mu_i = b'(\theta_i)$, $V(\mu_i) = b''(\theta_i)$

- where

$$\frac{\partial}{\partial \theta_i} \log L_i = \frac{\partial l_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}$$
$$\frac{\partial \theta_i}{\partial \mu_i} = 1 / \partial \mu_i / \partial \theta_i = 1 / b''(\theta_i)$$
$$\frac{\partial \mu_i}{\partial \eta_i} = 1 / \frac{\partial \eta_i}{\partial \mu_i} = 1 / g'(\mu_i)$$
$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}, \quad x_j = (x_{i1}, \ldots, x_{ip})^T$$

Let $v_i = a(\phi) V(\mu_i) = a(\phi) b''(\theta_i) = V(y_i)$

- Thus, we have

$$U_j(\beta) = \sum_i \frac{\partial l_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$
$$= \sum_i \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

* Use $\frac{\partial \mu_i}{\partial \eta_i} = 1 / \frac{\partial \eta_i}{\partial \mu_i} \longrightarrow$ $$= \sum_i \frac{1}{V_i (\frac{\partial \eta_i}{\partial \mu_i})^2} (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij}$$

Here we don't cancel $\frac{\partial \eta_i}{\partial \mu_i}$ to make IWLS form.

Estimation (continued)

→ • i.e., let $W_i = 1/\{V(Y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2\}$, with $V(Y_i) = \mathrm{Var}(Y_i) = a_i(\phi)V(\mu_i)$, then

$$U_j(\beta) \equiv \sum_{i=1}^{n} X_{ij} W_i (Z_i - \eta_i),$$

where

$$\begin{aligned} W_i^{-1} &= \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 a_i(\phi)V(\mu_i) = (g'(\mu_i))^2 a_i(\phi)V(\mu_i), \\ Z_i &= \eta_i + (Y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i} = \eta_i + (Y_i - \mu_i)g'(\mu_i) \\ &= \eta_i + \tilde{\varepsilon}_i. \end{aligned}$$

We see that $\mathrm{Var}(Z_i) \approx \mathrm{Var}(Y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 = 1/W_i = \mathrm{Var}(\tilde{\varepsilon}_i)$.

- Note: In CLMs, η_i is a linear model (i.e. $= \mathbf{X}_i^T\beta$). So the score function $U_j(\beta)$ above looks like a weighted score function of a linear regression model for Z_i with $W_i = 1/\mathrm{Var}(\tilde{\varepsilon}_i)$ In fact, since

$$Z_i = \eta_i + (Y_i - \mu_i)g'(\mu_i) \triangleq X_i^T\beta + \tilde{\varepsilon}_i$$
$$\mathrm{Var}(\tilde{\varepsilon}_i) = \mathrm{Var}(Y_i)(g'(\mu_i))^2 = 1/W_i,$$

then WLSE $\min_\beta \sum_i W_i (Y_i \quad X_i^T\beta)^2$

STAT 635-GLM-Lecture Notes 5, Estimation for Generalized Linear Models, Fall 2017

Estimation (continued)

- The score function in matrix notation:

$$U(\beta) = \mathbf{X}^T\mathbf{W}(\mathbf{Z} - \eta),$$

where

$$\begin{aligned} \mathbf{X}_{n\times p} &= (\mathbf{X}_1^T, \quad , \mathbf{X}_n^T)^T, \ \mathbf{W}_{n\times n} = \mathrm{diag}(W_1, \quad , W_n), \\ \mathbf{Z} &= (Z_1, \quad , Z_n), \ \eta = (\eta_1, \quad , \eta_n)^T = \mathbf{X}\beta. \end{aligned}$$

The score equation is

- Assuming full rank of $\mathbf{X}$, $U(\beta) = X^T W(Z - \eta) = X^T W(Z - X\beta) = 0$

$$U(\hat{\beta}) = 0 \quad \rightarrow \quad \hat{\beta} = (X^T W X)^{-1}(X^T W)Z$$

  - Problem: Z and W depend on β.

Estimation: Iterative WLS

- Iterative Weighted Least Squares (IWLS) algorithm for GLMs
  - Step 1. Initialization

$$\eta = g(\mathbf{Y}), \ \ (\mathbf{Y} \text{ initializes } \mu) \text{ or } \eta = \mathbf{X}\beta^{(\mathbf{0})},$$

← In Binomial and Poisson distributions to avoid $\log(0)$ for small values of Y in $\log(Y)$, use a small $\varepsilon > 0$, let $\eta = \log(\varepsilon)$

Note. $\eta = \log(Y)$ is a vector

  where $\beta^{(\mathbf{0})}$ is an initial value of β.

  Note: you may need some modification to avoid $\log(0)$.

  - Repeat the followings until changes in β are small (e.g., $\|\hat{\beta}^{(t+1)} - \hat{\beta}^{t}\| < 10^{-5}$ in SAS):

    (i) $\mu = g^{-1}(\eta)$.

    (ii) $\mathbf{Z} = \eta + (\mathbf{Y} - \mu)g'(\mu)$ (element-wise multiplication)

    (iii) $\mathbf{W} = \text{diag}(\{[g'(\mu)]^2 a(\phi)V(\mu)\}^{-1}$ (element-wise multiplication)

    (iv) $\beta = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Z}$.

    (v) $\eta = \mathbf{X}\beta$.

    (vi) Go back to (i).

IWLS (example)

- Independent r.v. $Y_1, \quad , Y_n \sim$ Poisson(λ_i) with p covariates X_i^T

$$f(y_i|\theta, \phi) = \frac{\lambda_i^{y_i} e^{\lambda_i}}{y_i!},$$

  with $\mu_i = \lambda_i$.

  - $\theta = \log(\lambda)$, $b(\theta) = e^{\theta}$ $a(\phi) = 1$.
  - Canonical link: $\eta = \theta = \log(\lambda) \to$ Log-linear regression

$$\log(\mu_i) = \mathbf{X}_i^T\beta,$$

- MLE of β For the Poisson model, $\theta_i = \log(\mu_i)$, $b(\theta_i) = e^{\theta_i}$, $a(\phi) = 1$, $\mu_i = b'(\theta_i)$

$$\frac{\partial}{\partial \theta_i} \log L_i = \frac{y_i - b'(\theta_i)}{a(\phi)} = y_i - e^{\theta_i}$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = \frac{1}{e^{\theta_i}}$$

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{\frac{\partial \eta_i}{\partial \mu_i}} = \frac{1}{g'(\mu_i)} = \frac{1}{1/\mu_i} = \mu_i = e^{\theta_i}$$

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$$

- The score function is

$$U_j(\beta) = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$
$$= (y_i - \mu_i) x_{ij}$$

In the ~~t~~form of IWLS, $u_j(\beta) = w_i(z_i - \eta_i) x_{ij}$,
where $w_i = \frac{1}{Var(y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} = \frac{1}{\mu_i \cdot (1/\mu_i)^2} = \mu_i = e^{\theta_i}$,
$z_i = x_i^T\beta + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + (y_i - \mu_i) \cdot \frac{1}{e^{\theta_i}}$

One can see that $Var(z_i) = Var(y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 = 1/w_i$

Using Proc/IML to implement IWLS

See Table4_3.sas for the plot in Figure 4.5, and Table4_4.sas for the results from IWLS in Table 4.4, P. 68

[In addition I have Table4_3 R to do the same calculation]

IWLS vs. Other Algorithms

Ex See data in Table 4.3 and Figure 4.5 in page 67 3rd edt
Poisson Regression with the identity link

- IWLS solves the score equations.
- There are other commonly used methods for solving score equations:
  - Fisher scoring:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + (\mathbf{I}^{(t)})^{-1} U^{(t)},$$

where

$$\mathbf{I}^{(t)} = -E\left[\frac{\partial^2 \log L}{\partial \beta \partial \beta^T}\right]_{\beta = \hat{\beta}^{(t)}},$$

the total information matrix for the sample.

  - Newton-Raphson:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^t + (\hat{\mathbf{I}}^{(t)})^{-1} U^{(t)},$$

where

$$\hat{\mathbf{I}}^{(t)} = -\frac{\partial^2 \log L}{\partial \beta \partial \beta^T}\Big|_{\beta = \hat{\beta}^{(t)}}$$

In this example, $g(\mu_i) = \eta_i \Rightarrow \mu_i = \eta_i$, then $\frac{\partial \eta_i}{\partial \mu_i} = 1$ and

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i} = \eta_i + (y_i - \mu_i) \cdot 1 = y_i$$

$$W_i = \frac{1}{Var(y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} = \frac{1}{Var(y_i)} = \frac{1}{\mu_i} = \frac{1}{\beta_1 + \beta_2 x_i}$$

The IWLS is $\mathcal{J}^{(m-1)} \mathbb{b}^{(m)} = \mathbb{X}^T \mathbb{W} \mathbb{Z}^{(m-1)}$

where the information matrix $\mathcal{J}$:

$$\mathcal{J} = \mathbb{X}^T \mathbb{W} \mathbb{X} = \begin{bmatrix} \sum_{i=1}^N \frac{1}{b_1 + b_2 x_i} & \sum_{i=1}^n \frac{x_i}{b_1 + b_2 x_i} \\ \sum_{i=1}^N \frac{x_i}{b_1 + b_2 x_i} & \sum_{i=1}^N \frac{x_i^2}{b_1 + b_2 x_i} \end{bmatrix}, \quad \mathbb{X}\mathbb{W}\mathbb{Z} = \begin{bmatrix} \sum_{i=1}^N \frac{y_i}{b_1 + b_2 x_i} \\ \sum_{i=1}^N \frac{x_i y_i}{b_1 + b_2 x_i} \end{bmatrix}$$