Lecture 31

Wednesday, April 3, 2024 10:47 AM

Agenda
- GLM

Def) Generalized linear Model (GLM)

A GLM has two components

1) Systematic component which relates the mean response to covariate values via a link function

$$E(Y_i) = \mu_i \qquad g(\mu_i) = \eta_i = X_i^T \beta$$

2) Random component which describes the distribution of the data. We will restrict ourselves to using distributions that are exponential dispersion families which have the form

$$f(y_i; \varphi_i, \theta) = e^{\frac{y_i\theta_i - b(\theta_i)}{\varphi_i} + c(y_i; \varphi_i)}$$

Properties of the exponential dispersion family

let $l = \ln(f(y_i; \varphi, \theta)) = \dfrac{Y_i\theta_i - b(\theta_i)}{\varphi_i} + c(y_i; \varphi_i)$ we will make use of two identities

1) $E\left(\frac{d}{d\theta_i}(l)\right) = 0$

2) $E\left(\frac{d^2}{d\theta_i^2}(l)\right) = -E\left(\left(\frac{d}{d\theta} l\right)^2\right)$

Lets look at $\dfrac{d}{d\theta_i} l = \dfrac{d}{d\theta_i}\left[\dfrac{Y_i\theta_i - b(\theta_i)}{\varphi_i} + c(y_i; \varphi_i)\right] = \dfrac{1}{\varphi_i}[Y_i - b'(\theta_i)]$

$$E\left(\frac{d}{d\theta_i} l\right) = 0 \implies E\left(\frac{1}{\varphi_i}[Y_i - b'(\theta_i)]\right) = 0 \implies E(Y_i) = b'(\theta_i) = \mu_i$$

Now, for $\dfrac{d^2}{d\theta_i^2} l = \dfrac{-b''(\theta_i)}{\varphi_i}$ so

$$E\left(\frac{d^2}{d\theta_i^2} l\right) = -E\left(\left(\frac{d}{d\theta_i} l\right)^2\right) \implies E\left(-\frac{b''(\theta_i)}{\varphi_i}\right) = -E\left(\frac{(Y_i - b'(\theta_i))^2}{\varphi_i^2}\right)$$

$$\implies \varphi_i\, b''(\theta_i) = E((Y_i - \mu_i)^2) = Var(Y_i)$$

Something to note, if $Var(Y_i) > 0$, then $b''(\theta_i) > 0$ so $b'(\theta_i)$ is monotonically increasing and thus has an inverse so

$$\mu_i = b'(\theta_i) \iff \theta_i = (b')^{-1}(\mu_i)$$

Now, from the systematic component, $g(\mu_i) = X_i^T\beta \implies \underline{\mu_i = g^{-1}(X_i^T\beta)}$. We call the canonical link the function g s.t.

$$\theta_i = X_i^T\beta \implies \theta_i = (b')^{-1}(g^{-1}(X_i^T\beta)) \implies g^{-1}(X_i^T\beta) = b'(\theta_i)$$

Using this, we can describe

$$Var(Y_i) = \varphi_i b''(\theta_i) = \varphi_i \underbrace{b''((b')^{-1}(\mu_i))} = \varphi_i v(\mu_i)$$

Estimating β:

Since $\mu_i = g(\eta_i) = g(X_i^T\beta)$ μ_i relates to β_i so we can find estimates for β using MLE

$$\frac{d}{d\beta_k} l = \frac{d}{d\beta_k}\sum\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i} + c(y_i;\varphi_i)\right] = \sum \frac{d}{d\beta_k}\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i} + c(y_i;\varphi_i)\right]$$

$$= \sum \frac{d}{d\beta_k}\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i}\right]$$

To find $\frac{d}{d\beta_k}\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i}\right] = \frac{d}{d\theta_i}\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i}\right]\frac{d\theta_i}{d\mu_i}\cdot\underbrace{\frac{d\mu_i}{d\eta_i}\cdot\frac{d\eta_i}{d\beta_k}}$

(a) $$\frac{d}{d\theta_i}\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi_i}\right] = \frac{y_i - b'(\theta_i)}{\varphi_i} = \frac{y_i - \mu_i}{\varphi_i}$$

(b) $$\frac{d\theta_i}{d\mu_i} = \frac{1}{\frac{d\mu_i}{d\theta_i}} = \frac{1}{b''(\theta_i)} = \frac{1}{b''((b')^{-1}(\mu_i))} = \frac{1}{v(\mu_i)}$$

$$\mu_i = b'(\theta_i) \qquad \frac{d\mu_i}{d\theta_i} = b''(\theta_i)$$

(c) $$\frac{d\mu_i}{d\eta_i} = \frac{1}{\frac{d\eta_i}{d\mu_i}} = \frac{1}{g'(\mu_i)}$$

$$g(\mu_i) = \eta_i \qquad \frac{d\eta_i}{d\mu_i} = g'(\mu_i)$$

(d) $$\frac{d\eta_i}{d\beta_k} = \frac{d}{d\beta_k}[x_i^T\beta] = \frac{d}{d\beta_k}(x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p) = x_{ik}$$

This tells us that

$$\frac{d}{d\beta_k} l = \sum \frac{y_i - \mu_i}{\varphi_i}\cdot\frac{1}{v(\mu_i)}\cdot\frac{1}{g'(\mu_i)}\cdot x_{ik} = \sum\frac{(y_i - \mu_i)x_{ik}}{\varphi_i v(\mu_i) g'(\mu_i)}$$

Newton's Method:

We can create a linear approximation of $f(x)$ about x_0 w/

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0)$$

If we want to find the value of x that makes $f(x)=0$, then

$$0 = f(x) \approx f(x_0) + (x-x_0) f'(x_0) \implies x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Algorithm:

1) initialize x_0
2) while $|f(x_0)| > \epsilon$

$$x_k = x_{k-1} - \frac{f(x_k)}{f'(x_k)}$$

Similarly, $f'(x) \approx f'(x_0) + (x-x_0) f''(x_0)$ which gives $x = x_0 - \dfrac{f'(x_0)}{f''(x_0)}$

In matrix form, newton's method looks like

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

Def/Information matrix

the observed information is $\hat{I} = \left[\dfrac{-d^2}{d\beta_k d\beta_j} \ell\right]_{k,j}$

the expected information is $I = E(\hat{I})$

So we need $\hat{I}$ for the GLM

$$\frac{d\ell}{d\beta_k d\beta_j} = \sum \frac{d}{d\beta_j}\left[\frac{(Y_i-\mu_i)x_{ik}}{\varphi_i V(\mu_i) g'(\mu_i)}\right]$$

$$= \sum \frac{d}{d\mu_i}\left[\frac{(Y_i-\mu_i)x_{ik}}{\varphi_i V(\mu_i) g'(\mu_i)}\right] \cdot \frac{d\mu_i}{d\beta_j}$$

$$= \sum \left[-\frac{\mu_i x_{ik}}{\varphi_i V(\mu_i) g'(\mu_i)} + \frac{(Y_i-\mu_i)x_{ik}}{\varphi_i} \underbrace{\frac{d}{d\mu_i}\left[\frac{1}{V(\mu_i) g'(\mu_i)}\right]}\right] \frac{x_{ij}}{g'(\mu_i)}$$

Now, finding $I = E(\hat{I})$, we get

$$I = E\left(\frac{-d\ell}{d\beta_k d\beta_j}\right) = \sum \frac{\mu_i x_{ik} x_{ij}}{\varphi_i V(\mu_i)[g'(\mu_i)]^2} + \cancel{\frac{x_{ij}}{\varphi_i g'(\mu_i)} \frac{d}{d\mu_i}\left[\frac{1}{V(\mu_i) g'(\mu_i)}\right] E(Y_i-\mu_i)} \to 0$$

$$= \sum \frac{\mu_i x_{ik} x_{ij}}{\varphi_i V(\mu_i)[g'(\mu_i)]^2}$$

We will write I and $\dfrac{d\ell}{d\beta}$ as a product of matrices. If we let $W = \operatorname{diag}\left\{\dfrac{1}{\varphi_i V(\mu_i)[g'(\mu_i)]^2}\right\}$, then

$$I = X^T W X$$

$I = X^T W X$

$\frac{dl}{d\beta} = X^T W u \qquad u_i = (Y_i - \mu_i) g'(\mu_i)$

By newton's method

$\hat{\beta}_{k+1} = \hat{\beta}_k + [I]^{-1} \nabla l \implies \hat{\beta}_{k+1} = \hat{\beta}_k + [X^T W X]^{-1} X^T W u$

$\implies X^T W X \hat{\beta}_{k+1} = X^T W X \hat{\beta}_k + X^T W u$

$= X^T W [X \hat{\beta}_k + u] \qquad Z = X \hat{\beta}_k + u$

$\implies X^T W X \hat{\beta}_{k+1} = X^T W Z$

$\implies \hat{\beta}_{k+1} = [X^T W X]^{-1} X^T W Z$

To find the MLE for β

<u>Algorithm:</u>

1) initialize β_0 and use that to calculate $\hat{\mu}_0$ and $\hat{W}_0, \hat{u}_0$

2) while $\|\hat{\beta}_{k+1} - \hat{\beta}_k\|^2 > \varepsilon$

   1) $\hat{\beta}_{k+1} = [X^T \hat{W}_k X]^{-1} X^T \hat{W}_k \hat{Z}_k$

   2) Find $\vec{\mu}_{k+1} = (b')^{-1} (X_i^T \hat{\beta}_{k+1})$