Generalized Linear Models

We have already discussed the exponential family

$$p_\theta(x) = \exp\{\eta(\theta)T(x) - A(\theta) + c(x)\}$$

$\longrightarrow$ now we will generalize to different means for each x_i via θ_i, $i=1, \dots, n$

Remember, each family has a mean variance relationship through $A(\theta)$

If $p_\theta(x)$ of form $\exp\{\theta x - A(\theta) + c(x)\}$

$$\Rightarrow \begin{cases} E(X) = \mu = A'(\theta) \\ V(X) = A''(\theta) = \dfrac{\partial E(X)}{\partial \theta} = v(\mu) \end{cases}$$

We will focus on the most commonly used type of model in applications — <u>exponential dispersion model</u> where we divide parameters into mean parameters of interest θ and nuisance dispersion parameter ϕ.

$$p_{\theta,\phi}(x) = \exp\left\{\frac{x\theta - A(\theta)}{\phi} + c(x,\phi)\right\}$$

Here $E(X) = \mu = A'(\theta)$

$$V(X) = \phi A''(\theta) = \phi v(\mu)$$

independent scale

how mean affects variance

Examples

Normal model $\quad p_{\mu,\sigma^2}(x) = \exp\left\{ \frac{\mu x - \frac{1}{2}\mu^2}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} \right\}$

$\Rightarrow \phi = \sigma^2 \quad , A(\mu) = \frac{1}{2}\mu^2$

$\Rightarrow \begin{cases} E(X) = A'(\mu) = \mu \\ V(X) = \phi A''(\mu) = \phi \end{cases}$

Poisson

$$p_\theta(x) = \frac{e^{-\theta}\theta^x}{x!} = \exp\{\log(\theta)x - \theta - \log x!\}$$

$\Rightarrow \boxed{\phi = 1} \quad \begin{cases} E(X) = \theta \\ V(X) = \theta \end{cases}$

Now consider dependent variable Y and independent variables $X_1, \dots, X_p$

$\Rightarrow$ observation "pairs" $\quad (\underset{\sim}{X}_i, y_i)_{i=1}^n \quad \underset{\sim}{X}_i = (x_{i1}, \dots, x_{ip})$

$$p_{\theta_i,\phi}(y_i | X_i) = \exp\left\{ \frac{y_i\theta_i - A(\theta_i)}{\phi} + c(y_i, \phi) \right\}$$

where $\begin{cases} E(y_i) = \mu_i = A'(\theta_i) \\ V(y_i) = \phi A''(\theta_i) = \phi v(\mu_i) \end{cases}$

and we assume $\begin{cases} h(\mu_i) = \underset{\sim}{x}_i'\beta = \eta_i \\ \mu_i = h^{-1}(\underset{\sim}{x}_i'\beta) \end{cases}$

i.e. mean μ_i is related to X through the <u>link function</u> $h()$ which is assumed known.

Now, since $\mu_i = A'(\theta_i) \Rightarrow$ means $g(\theta_i) = \underset{\sim}{X}_i'\beta$ for some function $g()$.

If $\theta_i = \underset{\sim}{x}_i' \beta$ we call the model form canonical

Problems addressed with GLM

First review other types of models

(1) Linear model

$y = X\beta + \varepsilon \quad E(\varepsilon) = 0$
$\qquad\qquad\qquad V(\varepsilon) = \sigma^2 I$ $\longrightarrow$ Least Squares

$\hat{\beta} = (X'X)^{-1}X'y$ closed form solution

$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

if assume $\varepsilon \sim N(0, \sigma^2)$ $\xrightarrow{MLE}$

(2) Nonlinear model

$y = f(X; \beta) + \varepsilon, \; E(\varepsilon) = 0$
$\qquad\qquad\qquad V(\varepsilon) = \sigma^2 I$ $\longrightarrow$ Least Squares $\longrightarrow$ no closed form solution usually

$\downarrow$

Solve by Iterative Least Squares

$\downarrow$

- Convergence can be a problem
- No closed form solution for $V(\hat{\beta})$.

if assume $\varepsilon \sim N(0, \sigma^2)$ $\xrightarrow{MLE}$

(3) Weighted Least Squares - heteroscedastic errors

$y_i = f(x_i; \beta) + \varepsilon_i$, $E(\varepsilon_i) = 0$
$V(\varepsilon_i) = \underline{\underline{\sigma_i^2}}$ $\longrightarrow$ Weighted Least squares

$$\downarrow$$

$$\min_{\beta} \sum_{i=1}^{n} w_i (y_i - \underset{\sim}{x}_i' \beta)^2$$

$$= \min_{\beta} \| W^{1/2} y - W^{1/2} X \|^2$$

where W usually $\text{diag}\left(\frac{1}{\sigma_i^2}\right)$

$$\begin{cases} \hat{\beta} = (X'WX)^{-1} X'Wy \\ V(\hat{\beta}) = \sigma^2 (X'WX)^{-1} \end{cases}$$

$\left(\text{where we assume } \begin{cases} V(y_i) = \sigma^2 \left(\frac{\sigma_i^2}{\sigma^2}\right) \\ w_i = \frac{\sigma^2}{\sigma_i^2} \end{cases} \right)$

relative variance

(4) GLM

$y \sim f_y(x; \beta)$ — allow for different error structure, e.g. poisson model

— allow for variance-mean dependence

- Assume independent observations y_i
- Assume $E(y_i) = \mu_i$, $h(\mu_i) = \underset{\sim}{x}_i' \beta$ for <u>known</u> $h()$
  $= \eta_i$
  "linear predictor"
- Assume known mean-variance relationship $V(y_i) = \phi \, v(\mu_i)$

Stay with exponential dispersion family

(5)

$$P_{\theta_i,\phi}(y_i|\underset{\sim}{x}_i) = \exp\left\{\frac{y_i\theta_i(x_i) - A(\theta_i)}{\phi} + c(y_i,\phi)\right\}$$

where $\begin{cases} E(y_i) = \mu_i = A'(\theta_i(x_i)) \Rightarrow h(\mu_i) = \underset{\sim}{x}_i'\beta = \eta_i = g(\theta_i) \\ V(y_i) = \phi v(\mu_i) \end{cases}$

How do we estimate β's? How do inference?

In general, we do not get a closed form solution $\hat{\beta}$.

<u>MLE</u>

$$\log L(\underset{\sim}{\theta},\phi|y) = \ell(\underset{\sim}{\theta},\phi|y) = \sum_{i=1}^{n} \frac{y_i\theta_i - A(\theta_i)}{\phi} + c(y_i,\phi) \qquad \left[g(\theta) = \underset{\sim}{x}_i'\beta\right]$$

$$\Rightarrow \ell(\beta,\phi|y) = \sum_{i=1}^{n} \frac{y_i g^{-1}(\underset{\sim}{x}_i'\beta) - A(g^{-1}(\underset{\sim}{x}_i'\beta))}{\phi} + c(y_i,\phi)$$

- Let's maximize ℓ wrt one β at a time
- Also note, $\ell(\beta,\phi|y) = \sum_{i=1}^{n} \underbrace{\ell(\beta,\phi|y_i)} = \sum_{i=1}^{n} \ell_i$

look @ $\frac{\partial \ell_i}{\partial \beta_j}$ and understand its form

$\Rightarrow$ $\ell()$ is additive so can look at form of $\frac{\partial \ell}{\partial \beta}$ one at a time

$$\boxed{\frac{\partial \ell}{\partial \beta_j} = \frac{\partial \ell}{\partial \theta}\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta_j} = 0}$$

MLE solution at zero-crossing of $\frac{\partial \ell}{\partial \beta_j}$

OK, so we want β's such that

$$\boxed{\frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \theta}\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta_j} = 0}$$ for all j

- $\eta = \underset{\sim}{X}'\beta \Rightarrow \frac{\partial \eta}{\partial \beta_j} = X_j$

- $l(\theta) = \frac{y\theta - A(\theta)}{\phi} + c(y,\phi) \Rightarrow \frac{\partial l}{\partial \theta} = \frac{y - A'(\theta)}{\phi} = \frac{y-\mu}{\phi}$

- $\mu = A'(\theta)$

$\frac{\partial \mu}{\partial \theta} = A''(\theta) = V$ where $V(y) = \phi \cdot V$ $\Rightarrow \frac{\partial \theta}{\partial \mu} = \frac{1}{V}$

$$\Longrightarrow \frac{\partial l}{\partial \beta_j} = \frac{y-\mu}{\phi}\cdot\frac{1}{V}\cdot\frac{\partial \mu}{\partial \eta}\cdot X_j$$

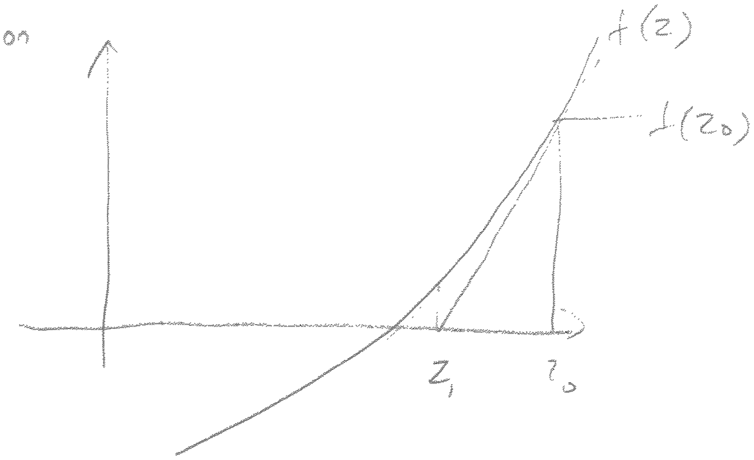
$\left[\text{Define } W^{-1} = \left(\frac{\partial \eta}{\partial \mu}\right)^2 V\right]$

$$= \frac{W}{\phi}(y-\mu)\frac{\partial \eta}{\partial \mu}X_j$$

Need to solve $\frac{W}{\phi}(y-\mu)\frac{\partial \eta}{\partial \mu}X_j = 0$ for all j

How do we find zero-crossings?

Newton-Rhapson



Finding zero-crossings of $f(z)$

(1) Take a current value z_0 and linearize $f(z)$ at z_0

Line through $f(z_0)$ has form $a + f'(z_0)\, z_0 = f(z_0)$

$\Rightarrow$ solve for a: line has form $\underline{a + f'(z_0) \cdot z}$

(2) Extend to the zero-crossing point z_1

where $a + f'(z_0) z_1 = 0$

$$\Rightarrow \boxed{z_1 = z_0 - \underbrace{\frac{f(z_0)}{f'(z_0)}}_{\text{stepsize}}}$$

update

(3) Iterate

Now, we will apply Newton-Rhapson to $\frac{\partial l}{\partial \beta} = f(\beta)$

Initial value for β is β_0

Iterative Weighted Least Squares

- We have $E(y_i) = \mu_i \quad , h(\mu_i) = \underset{\sim}{x_i}'\beta = \eta_i$

- Taylor approximation of $h(\mu)$ around μ_0

$$h(\mu) \approx h(\mu_0) + (\mu - \mu_0) \left.\frac{\partial h(\mu)}{\partial \mu}\right|_{\mu_0}$$

$$= \eta_0 + (\mu - \mu_0) \left.\frac{\partial \eta}{\partial \mu}\right|_{\mu_0}$$

$$= x'\beta_0 + (\mu - \mu_0) \left.\frac{\partial \eta}{\partial \mu}\right|_{\mu_0} \qquad \text{where } \mu_0 = h^{-1}(x'\beta_0)$$

- Define <u>working values</u>

$$\boxed{Z_0 = \eta_0 + (y - \mu_0) \left.\frac{\partial \eta}{\partial \mu}\right|_{\mu_0}}$$

"linearization of the data"

y instead of unknown true mean μ.

<u>working weights</u>

$$W_0^{-1} = \left(\frac{\partial \eta}{\partial \mu}\right)^2_{\mu_0} V_0 \quad , \text{ where } V_0 = A''(\theta)\Big|_{\theta_0} = \left.\frac{\partial \mu}{\partial \theta}\right|_{\theta_0}$$

= "Variance of Z_0" since $V(y) = \phi A''(\theta)$

- Update β by regressing Z_0 on X with weights W_0! Iterate

$\Rightarrow$ Need to show this is equivalent to MLE estimation of β

At starting point β_0 we have

$$f_j(\beta_0) = W_0 (y-\mu_0) \left.\frac{\partial \eta}{\partial \mu}\right|_{\mu_0} x_j$$

For all components of β (all j) we have

$$\boxed{f(\beta_0) = X' W_0 (z_0 - \eta_0)} \qquad \text{since } z_0 = \eta_0 + \left.\frac{\partial \eta}{\partial \mu}\right|_{\mu_0} (y-\mu_0)$$

$\eta_0 = X'\beta_0$

all depend on β_0

To update β : Newton-Rhapson $\quad \beta_{new} = \beta_0 - \dfrac{f(\beta_0)}{f'(\beta_0)}$

Here $\quad f'(\beta_0) = [j,j' \text{ components}] = \left.\dfrac{\partial^2 l}{\partial \beta_j \partial \beta_{j'}}\right|_{\beta_0}$

$$= \frac{\partial}{\partial \beta_{j'}} \left.\left(\frac{W}{\phi} (y-\mu) \frac{\partial \eta}{\partial \mu} x_j \right)\right|_{\beta_0}$$

$$= \left(\underbrace{(y-\mu) \frac{\partial}{\partial \beta_{j'}} \left.\left(\frac{W}{\phi} \frac{\partial \eta}{\partial \mu} x_j \right)\right|_{\beta_0}}_{1} + \underbrace{W \frac{\partial \eta}{\partial \mu} x_j \frac{\partial}{\partial \beta_{j'}} \left.(y-\mu)\right|_{\beta_0}}_{2} \right)$$

Trick called <u>Fisher scoring</u> : replace $f'(\beta_0)$ by its <u>expectation</u>.

$\Rightarrow$ then term 1 is 0!

$\Rightarrow$ term 2 $\quad W \dfrac{\partial \eta}{\partial \mu} x_j \left(-\dfrac{\partial \mu}{\partial \eta} \dfrac{\partial \eta}{\partial \beta_{j'}} \right) = -W x_j x_{j'}$

for all j, j' pairs

$$\nabla^2 \ell = \left.\frac{\partial^2 \ell}{\partial\beta_j \partial\beta_{j'}}\right|_{\substack{\text{all} \\ j,j'}} \Bigg|_{\beta_0} = \underline{\underline{-X' W_0 X}}$$

$$\Rightarrow \beta_{new} = \beta_0 - E\left(f'(\beta_0)\right)^{-1} f(\beta_0)$$

$$= \beta_0 + (X' W_0 X)^{-1} X' W_0 (Z_0 - \underbrace{\eta_0}_{= X\beta_0})$$

$$= (X' W_0 X)^{-1} X' W_0 Z_0$$

= weighted least squares solution

regressing Z_0 on X with weights W_0

<u>So</u> MLE $\overset{\text{equivalent}}{\underset{\text{to}}{\Longleftrightarrow}}$ iterative weighted least squares $\Rightarrow$ final $\hat\beta$

We can also approximate $V(\hat\beta)$ by the last

final weighted least squares

$$V(\hat\beta) = \phi (X' W_{final} X)^{-1}$$

Inference

CI for β: common to use normal approximation

$$\left[\hat{\beta}_j \pm z_{1-\alpha/2}\sqrt{V(\hat{\beta}_j)}\right] \quad \text{with} \quad V(\hat{\beta}_j) = \phi(X'W_{final}X)^{-1}_{jj}$$

but can be a poor approximation if

a) local linearization not adequate near $\hat{\beta}$

b) $\phi_{real} >> \phi_{assumed}$ (overdispersion)

c) normal distribution far from actual sampling distribution

Analysis of Deviance

- Deviance of model M: $D(M) = -2 * \text{loglike}(\hat{\beta}_M)$
- Scaled deviance $\frac{D(M)}{\phi}$
- If M is an adequate model, then for large n

$$\frac{D(M)}{\phi} \sim \chi^2_{n-p}$$

If $\frac{D(M)}{\phi} >> \chi^2_{n-p}(1-\alpha) \rightarrow$ reject model M

"goodness-of-fit" test

Overdispersion

$\frac{D(M)}{\phi} \sim \chi^2_{n-p}$ if take expectation $\Rightarrow E\left(\frac{D(M)}{\phi}\right) = n-p$

$\Rightarrow \hat{\phi} = \frac{D(M)}{n-p}$ [example normal model $\hat{\sigma}^2 = \frac{RSS}{n-p}$]

if $\hat{\phi} >> \phi_{assumed}$ we say we have overdispersion

$\Rightarrow$ Can use $\hat{\phi}$ to inflate estimate of $V(\hat{\beta})$

$\Rightarrow V(\hat{\beta}) = \hat{\phi}(X'W_{final}X)^{-1}$ $\Rightarrow$ use this for CI

If $\phi_{real} >> \phi_{assumed}$, not using $\hat{\phi}$ can lead to CIs that are much too narrow

Comparing models

- One alternative is to use $AIC = D(M) + 2p_M$ to choose between models.
- Alternative 2
- M_1 model with p_1 parameters nested under M_2 with $p_2 = p_1 + \Delta p$ parameters

$\Rightarrow D(M_1) - D(M_2) \sim \chi^2_{\Delta p}$ under null M_1

Reject M_1 if $D(M_1) - D(M_2) > \chi^2_{\Delta p}(1-\alpha)$

if $\hat{\phi} >> \phi_{assumed}$ can use ad-hoc F-test

$\frac{D(M_1) - D(M_2)}{\hat{\phi}(p_2 - p_1)} \sim F_{p_2-p_1,\, n-p_2}$ under null M_1