538 Lec. #3

GLM

$\uparrow$

- How G?
- What is "under the hood"?

Prelim: Exponential family

$$f(y|\theta,\phi) = \exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi) \right\}$$

↑ density or pmf

Moments: $E(Y) = b'(\theta)$

$Var(Y) = a(\phi)\, b''(\theta)$

Defines variance function $v(\ )$ such that

$$Var(Y) = a(\phi)\, v\{E(Y)\}$$

$N(\mu,\sigma^2)$ $\theta = \mu$ $b(\theta) = \frac{1}{2}\theta^2$ $a(\phi) = \sigma^2$ $v(\mu) = 1$

Bernoulli(μ) $\theta = \log(\mu/1-\mu)$ $b(\theta) = \log(1+e^{\theta})$ $a(\phi) = 1$

check

$v(\mu) = \mu(1-\mu)$

Poisson(μ) $\theta = \log\mu$ $b(\theta) = e^{\theta}$ $a(\phi) = 1$ $v(\mu) = \mu$

Prelim: link function ← modelling choice

Assume $\underset{\sim}{X}$ influences Y via

linear predictor $\eta = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1}$

How are η and $\mu = E(Y)$ related?

Choose link function $g(\,)$ such that

$\in (-\infty, \infty)$ → $\eta = g(\mu)$

↑ range depends on data type

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

e.g. logistic regression

$g(\mu) = \log(\mu/1-\mu)$

LM $g(\mu) = \mu$

Canonical link?

(equates η and θ

e.g. binary Y

$g(\mu) = \log(\mu/1-\mu)$ is canonical

$g(\mu) = \log(-\log(1-\mu))$ is not

Prelim: Fisher Scoring $\ell(\beta)$ is log-likelihood   4

Applying Newton-Raphson as iterative scheme to solve $\boxed{\ell'(\underset{\sim}{\beta}) = \underset{\sim}{0}}$

$$\hat{\underset{\sim}{\beta}}^{next} = \hat{\underset{\sim}{\beta}}^{prev} + \left[-\ell''(\hat{\underset{\sim}{\beta}}^{prev})\right]^{-1} \ell'(\hat{\underset{\sim}{\beta}}^{prev})$$

FISHER SCORING MODIFICATION?

replace $-\ell''(\hat{\beta}^{prev})$ with $I(\hat{\underset{\sim}{\beta}}^{prev}) = E\left\{-\ell''(\hat{\beta}^{prev})\right\}$

w.r.t. $Y \mid \underset{\sim}{X}$

$$I(\beta) = \sum_{i=1}^{n} E\left\{-\frac{\partial^2}{\partial\beta^2} \log f(y_i \mid x_i; \beta)\right\}$$

wrt $f(y_i \mid x_i; \beta)$

WHY? — stable
— pos. def. everywhere

Prelim / Reminder: Weighted LS

$$\hat{\underset{\sim}{\beta}} = \underset{\underset{\sim}{\beta}}{\operatorname{argmin}} \left(\underset{\sim}{y} - X\underset{\sim}{\beta}\right)^T W \left(\underset{\sim}{y} - X\underset{\sim}{\beta}\right)$$

$$= \left(X^T W X\right)^{-1} X^T W \underset{\sim}{y}$$

e.g. arises if $Y_i \sim N(\underset{\sim}{x}_i^T \underset{\sim}{\beta}, \sigma_i^2)$

with $$W = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n^2 \end{pmatrix}$$

NOTE: W only matters up to constant of proportionality

Assembling pieces

A $$\frac{\partial}{\partial \underset{\sim}{\beta}} \log f(y_i ; \underset{\sim}{\beta}) = \frac{y_i - \mu_i}{a(\phi)\, v(\mu_i) \underbrace{\partial \eta_i / \partial \mu_i}_{= g'(\mu)}} \underset{\sim}{x}_i$$

B $$E\left\{ - \frac{\partial^2}{\partial \underset{\sim}{\beta}^2} \log f(y_i ; \underset{\sim}{\beta}) \right\} = \frac{1}{a(\phi)\, v(\mu_i) (\partial \eta_i / \partial \mu_i)^2} \underset{\sim}{x}_i \underset{\sim}{x}_i^T$$

C $$E\{ - \ell''(\underset{\sim}{\beta}) \} = \frac{1}{a(\phi)} \underset{p \times n}{X^T} W_\beta \underset{n \times p}{X}$$

where $$(W_\beta)_{ii} = \frac{1}{v(\mu_i) (\partial \eta_i / \partial \mu_i)^2}$$

D $$\ell'(\underset{\sim}{\beta}) = \frac{1}{a(\phi)} X^T W_\beta \begin{pmatrix} \partial \eta_1 / \partial \mu_1 \, (y_1 - \mu_1) \\ \vdots \\ \partial \eta_n / \partial \mu_n \, (y_n - \mu_n) \end{pmatrix}$$

So Fisher scoring algorithm is

$$\hat{\beta}^{(t+1)} = \beta^{(t)} + \left\{X^T W_{\hat{\beta}^{(t)}} X\right\}^{-1} X^T W_{\hat{\beta}^{(t)}} \underbrace{\frac{\partial \eta}{\partial \mu}(Y-\mu)}$$

$$= \left\{X^T W_{(t)} X\right\}^{-1} X^T W_{(t)} Y^*_{(t)}$$

where

$$Y^*_{(t)} = X \overset{\text{fitted}}{\hat{\beta}^{(t)}} + \overbrace{\frac{\partial \eta}{\partial \mu}(Y - \mu)}^{\text{residual}}$$

pseudo-response on linear predictor scale

converts to linear predictor scale

depend on $\hat{\beta}^{(t)}$

Algorithm

- for specified $g(\ )$, $v(\ )$ (link, variance)
- inputted X, Y

current $\hat{\beta}$ $\longrightarrow$ current $\hat{\eta} = X\hat{\beta}$

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

$\hat{v} = v(\hat{\mu})$

weights $= \dfrac{1}{\hat{v}\,(g'(\hat{\mu}))^2}$

pseudo-responses $\hat{\eta} + g'(\hat{\mu})(y-\hat{\mu})$

weighted LS

new $\hat{\beta}$

ex p. 118-119 fitting logistic regression use 5 iterations of lm( ) get same $\hat{\beta}$ as glm()