Lecture for STOR 556   3/19/19

Follow Chapter 8 (and a bit of Chapter 9) from the course text.

Definition of a GLM

Start with an exponential family

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

y: observed value of a RV

θ: canonical parameter

ϕ: dispersion parameter (scale)

a, b, c: known functions. What these functions are defines the family.

Ex. 1 Normal

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

$$= \exp\left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

Key term is $\frac{y\mu}{\sigma^2} = \frac{y\theta}{\phi}$ if we define $\theta = \mu$, $\phi = \sigma^2$

Then $f(y;\theta,\phi) = \exp\left[\frac{y\theta - \theta^2/2}{\phi} - \frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)\right]$

so $a(\phi) = \phi$, $b(\theta) = \frac{\theta^2}{2}$, $c(y,\phi) = -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$

3/19/19 p.2

Ex 2 Poisson

$\cancel{f(y;\mu) = y^{\mu}}$

$$f(y;\mu) = \frac{\mu^y e^{-\mu}}{y!} = \exp\left[y\log\mu - \mu - \log(y!) \right]$$

$$= \exp\left[y\theta - e^{\theta} - \log(y!) \right]$$

Here $\phi = 1$, $\theta = \log\mu$, $b(\theta) = e^{\theta}$, $c(y,\phi) = -\log y!$

Ex 3 Binomial – treat n as known

$$f(y;\mu) = \binom{n}{y} \mu^y (1-\mu)^{n-y}$$

$$= \exp\left[y \log\frac{\mu}{1-\mu} + n\log(1-\mu) + \log\binom{n}{y} \right]$$

Write $\theta = \log\frac{\mu}{1-\mu}$, $\mu = \frac{e^{\theta}}{1+e^{\theta}}$, $\log(1-\mu) = -\log(1+e^{\theta})$

$$f(y;\theta) = \exp\left[y\theta - n\log(1+e^{\theta}) + \log\binom{n}{y} \right]$$

so $\phi = 1$, $\theta = \log\frac{\mu}{1-\mu}$, $b(\theta) = +n\log(1+e^{\theta})$, $c(y,\phi) = \log\binom{n}{y}$.

Ex 4 Gamma distribution (p.175)

Usually written $\frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$, Faraway writes this as

$$\frac{\lambda^{\nu}}{\Gamma(\nu)} y^{\nu-1} e^{-\lambda y}$$

The trouble is, if we just write $\theta = -\lambda$, so $-\lambda y = +y\theta$, there's no place for ν to be recorded as a function of λ.

Faraway write $\lambda = \frac{\nu}{\mu}$ and recodes in terms of μ, ν:

$$f(y;\mu,\nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1} e^{-\frac{y\nu}{\mu}}$$

Define $\theta = -\frac{1}{\mu}$, $\phi = \frac{1}{\nu}$

$$f(y;\theta,\mu) = \exp\left[\frac{y\theta}{\phi} - \frac{1}{\phi}\log\left(-\frac{1}{\theta}\right) + \left(\frac{1}{\phi}-1\right)\log y - \frac{1}{\phi}\log\phi - \log\Gamma\left(\frac{1}{\phi}\right)\right]$$

$b(\theta) = \log\left(-\frac{1}{\theta}\right) = \log(-\theta)$ (Here $\theta < 0$)

Ex 5 Inverse ~~Gamma~~ Gaussian $IG(\mu, \lambda)$ (p.181)

$$f(y;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} \cdot \exp\left[-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right], \quad y,\mu,\lambda > 0.$$

Exercise Show that this is a GLM and hence (using results below) derive the formulas at the bottom of p. 181:

Mean $= \mu$, variance $= \frac{\mu^3}{\lambda}$

** typo in book

Mean and Variance

$$f(y;\theta,\varphi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\varphi)\right].$$

Log Likelihood for a single y (treat φ as known):

$$\ell(\theta) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)$$

$$\ell'(\theta) = \frac{y - b'(\theta)}{a(\phi)}$$

Likelihood theory implies $E(\ell'(\theta)) = \frac{Y - b'(\theta)}{a(\phi)} = 0$

Therefore, $EY = b'(\theta)$

Continue, $\ell''(\theta) = \frac{-b''(\theta)}{a(\phi)}$

Likelihood theory implies $-E(\ell''(\theta)) = E(\ell'(\theta))^2$

so $\frac{b''(\theta)}{a(\varphi)} = E\left(\frac{Y - b'(\theta)}{a(\phi)}\right)^2$, $\operatorname{var} Y = b''(\theta)a(\phi)$.

Also write as $\operatorname{var} Y = \frac{\phi b''(\theta)}{w}$ where $w = \frac{\phi}{a(\phi)}$ is a weight

Henceforth write $\mu = b'(\theta)$.

Ex1 $b(\theta) = \frac{\theta^2}{2}$ $b'(\theta) = \theta$ $b''(\theta) = 1$

so Mean $= \theta = \mu$, Var $= a(\phi) = \phi = \sigma^2$

Ex2 $b(\theta) = e^\theta \quad b'(\theta) = e^\theta = \mu$

$b''(\theta) = e^\theta = \mu$

mean $= \mu \qquad$ var $= \mu$

Ex3 $b(\theta) = -n\log(1+e^\theta) \quad b'(\theta) = +n\dfrac{e^\theta}{1+e^\theta} = n\left(1 - \dfrac{1}{1+e^\theta}\right) = n\mu$

$b'(\theta) = n\left(1 - \dfrac{1}{1+e^\theta}\right) \qquad b''(\theta) = \dfrac{ne^\theta}{(1+e^\theta)^2} = n \cdot \dfrac{e^\theta}{1+e^\theta} \cdot \dfrac{1}{1+e^\theta}$

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

Ex4, Ex 5, check for yourselves

Link functions Observation y_i has mean $\mu_i = b'(\theta_i)$ which depends on covariate $x_{i1}, \ldots, x_{ip}$.

Assume there is some function g for which

$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$

g is called the link function.

g is arbitrary and in some models there are several standard link functions, e.g. for binomial regression, we can have logit link, probit link, complementary log-log link, and others.

However there is one natural link called the canonical link function and this is often taken as the default. This is when $g(\mu) = \theta$

Since $\mu(\theta) = b'(\theta)$, needs $g(b'(\theta)) = \theta$.

e.g. Normal: $\mu = \theta$ so $g(\mu) = \mu$

Poisson $\mu = e^{\theta}$ so $\theta = \log \mu$

Binomial $\theta = \log \frac{\mu}{1-\mu}$. Note in this case the mean is actually $n\mu$, not μ, but it doesn't matter because it makes no difference which one is expressed as a linear function of covariates.

Gamma Literal application would lead to $\eta = -\frac{1}{\mu}$ but in practice we use $\eta = +\frac{1}{\mu}$ because that amounts to the same thing.

Fitting a GLM

Start by writing log likelihood where we assume $a_i(\phi) = \frac{\phi}{w_i}$

$$\ell(\beta; y_i) = \frac{w_i}{\phi}\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)$$

Solve $\sum_i \frac{\partial \ell(\beta; y_i)}{\partial \beta_j} = 0$

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi}\sum_i w_i\left(y_i \frac{\partial \theta_i}{\partial \beta_j} - b'(\theta_i)\frac{\partial \theta_i}{\partial \beta_j}\right).$$

Chain rule: $\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\{b'(\theta_i)\} = b''(\theta_i)\frac{\partial \theta_i}{\partial \beta_j}$

so $$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi}\sum_i \frac{y_i - b'(\theta_i)}{b''(\theta_i)/w_i} \cdot \frac{\partial \mu_i}{\partial \beta_j}$$

$$= \sum_i \frac{y_i - b'(\theta_i)}{V(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \beta_j} = \sum_i \frac{y_i - \mu_i}{V(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \beta_j}$$

where $V(\mu_i)$ is variance function

Algorithm (Faraway p.155)

I'll denote the iteration number by k instead of i.

Start with $k=0$

Initial estimates $\hat{\beta}^{(0)}$ leads to $\hat{\eta}^{(0)}$ and hence $\hat{\mu}^{(0)}$.

1. Form "adjusted variable" $z^{(k)} = \hat{\eta}^{(k)} + (y - \hat{\mu}^{(k)}) \left.\frac{\partial \eta}{\partial \mu}\right|_{\hat{\eta}^{(k)}}$

2. New weights $\frac{1}{w^{(k)}} = \left(\frac{\partial \eta}{\partial \mu}\right)^2_{\hat{\eta}^{(k)}} V(\hat{\mu}^{(k)})$

3. Regress $z^{(k)}$ on X with weights $w^{(k)}$
 $\Rightarrow$ new estimates $\hat{\beta}^{(k+1)}$

4. Set $k=k+1$ and return to step 1.

Bliss example (p.155) Retool binomial distribution ~~using~~ defining y_i as the proportion of successes in sample i of size n_i

$$f(y_i; \mu_i) = \binom{n_i}{n_i y_i} \mu_i^{n_i y_i} (1-\mu_i)^{n_i - n_i y_i}$$

$$= \exp\left[n_i y_i \log \frac{\mu_i}{1-\mu_i} + n_i \log(1-\mu_i) + \log \binom{n_i}{n_i y_i}\right]$$

$$= \exp\left[n_i y_i \theta_i - n_i \log(1+e^{\theta_i}) + \log\binom{n_i}{n_i y_i}\right]$$

$\eta_i = \theta_i = \log \frac{\mu_i}{1-\mu_i}$ $\quad \frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i(1-\mu_i)}$ $\quad V(\mu_i) = \frac{\mu_i(1-\mu_i)}{n_i}$

$\frac{1}{w_i} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 \cdot V(\mu_i) = \frac{1}{n_i \mu_i (1-\mu_i)}$ so $w_i = n_i \mu_i (1-\mu_i)$