Lecture 16

Recap: quasi-likelihood methods

usual GLM assumption:

$$Y_i \sim EDM(\mu_i, \phi)$$

$$\Rightarrow \mathbb{E}[Y_i] = \mu_i \quad Var(Y_i) = \phi V(\mu_i)$$

$$g(\mu_i) = \beta^T X_i$$

Quasi-likelihood method:

assume Y_i has some (unknown) distribution, with

$$\mathbb{E}[Y_i] = \mu_i = g^{-1}(\beta^T X_i)$$

$$Var(Y_i) = \phi V(\mu_i)$$

- describe relationship between X_i & μ_i
- describe relationship between μ_i & $Var(Y_i)$

Recap: quasi-Poisson

Poisson: $Y_i \sim \text{Poisson}(\mu_i)$ $\Rightarrow$ $\mathbb{E}[Y_i] = \mu_i$, $\text{Var}(Y_i) = \mu_i$

$(\phi = 1)$

$\log(\mu_i) = \beta^T X_i$

Quasi-Poisson: Y_i has <u>some</u> distribution with

$\mathbb{E}[Y_i] = \mu_i = \exp\{\beta^T X_i\}$

$\text{Var}(Y_i) = \phi \, \mu_i$

↑ allowing variance > mean (or <)

$\Rightarrow$ $\hat{\beta}_{QP} = \hat{\beta}_{\text{Poisson}}$ (estimating same mean function!)

$\text{Var}(\hat{\beta}_{QP}) = \phi \, \text{Var}(\hat{\beta}_{\text{Poisson}})$

↑ inflate variance to correct for overdispersion

When to use quasi-Poisson models

- The response is a count variable
- The Poisson shape assumption looks good
- The response has more variability than a Poisson distribution accounts for
- We believe variance is a multiple of the mean

Estimating the dispersion parameter

Motivation: if $Y_i \sim N(\mu_i, \sigma^2)$ $\quad (V(\mu_i) = 1)$

then $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2$

Two views:

$= \frac{D(Y, \hat{\mu})}{n-p}$ $\qquad$ $\frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{Y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \right)^2$

(mean deviance estimate) $\qquad$ (Pearson estimate)

In general: $\hat{\phi}_D = \frac{D(Y, \hat{\mu})}{n-p}$

$\hat{\phi}_P = \frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{Y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \right)^2$

$\hat{\phi}_D \neq \hat{\phi}_P$ in general (Normal is the only exception)

Analogous to $\hat{\sigma}^2 = \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$

For Poisson: saddlepoint approx. is good if $\min\{Y_i\} \geq 3$

Mean deviance estimate

$$f(y;\mu,\phi) = b(y,\phi)\exp\left\{-\frac{d(y,\mu)}{2\phi}\right\}$$

(saddlepoint approx)

$$\approx \frac{1}{\sqrt{2\pi\phi V(y)}}\exp\left\{-\frac{d(y,\mu)}{2\phi}\right\}$$

<u>If</u> saddlepoint approx. is good:

$$\frac{D(Y,\hat{\mu})}{\phi} \approx \chi^2_{n-p} \Rightarrow \mathbb{E}\left[\frac{D(Y,\hat{\mu})}{n-p}\right] \approx \phi$$

<u>Also</u>: $\ell(\mu,\phi) \approx \sum_{i=1}^{n}\left\{-\frac{1}{2}\log(2\pi\phi V(y)) - \frac{1}{2\phi}d(Y_i,\mu_i)\right\}$

$$\Rightarrow \frac{\partial \ell(\mu,\phi)}{\partial\phi} = \sum_{i=1}^{n}\left\{-\frac{1}{2}\cdot\frac{1}{\phi} + \frac{1}{2\phi^2}d(Y_i,\mu_i)\right\} \overset{\text{set}}{=} 0$$

$$\hat{\phi}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} d(Y_i,\mu_i) = \frac{D(Y,\mu)}{n}$$

Don't know μ, so plug $\hat{\mu}$: $\hat{\phi}_D = \frac{D(Y,\hat{\mu})}{n-p}$

Pearson estimate

$$\mathbb{E}[Y_i] = \mu_i \quad \Rightarrow \quad \mathbb{E}\left[\frac{Y_i - \mu_i}{\sqrt{V(\mu_i)}}\right] = 0$$

$$\mathbb{E}\left[\left(\frac{Y_i - \mu_i}{\sqrt{V(\mu_i)}}\right)^2\right] = \frac{\mathbb{E}[(Y_i - \mu_i)^2]}{V(\mu_i)} = \frac{\mathrm{Var}(Y_i)}{V(\mu_i)} = \frac{\phi V(\mu_i)}{V(\mu_i)} = \phi$$

$$\hat{\phi}_p = \frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{Y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}\right)^2$$

If appropriate assumptions hold:

$\hat{\phi}_p, \hat{\phi}_D \approx \phi$ (both approximately unbiased)

Distribution of the estimates

Linear regression: $\frac{(n-p)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$

GLMs: $\frac{(n-p)\,\hat{\phi}}{\phi} \approx \chi^2_{n-p}$

Inference with quasi-Poisson models

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_16.html