# STAT 770 Oct. 5 Lectures

# GLMs – Information & Estimating Equations

Reading and Topics for this lecture: Chapter 4 through Sec. 4.3

**(1)** Score Estimating Eq'n, Observed Information

**(2)** Logistic Regression

**(3)** Poisson Regression

**(4)** GLM as generalization: Exponential Families

**(5)** `R` coding – Function `glm`

# Score Equation – Observed Information

Model $f(x_i, \theta)$, iid data $X_1, \ldots, X_n$, $\log L(\theta) = \sum_{i=1}^n \log f(X_i, \theta)$

Calculus-maximizer $\widehat{\theta}$ MLE satisfies the **score equation**:

$$\nabla_\theta \log L(\theta) = \begin{pmatrix} \partial/\partial\theta_1 \\ \vdots \\ \partial/\partial\theta_p \end{pmatrix} = \sum_{i=1}^n \nabla_\theta \log f(X_i, \theta) = \mathbf{0}$$

Taylor expansion (Mean Value Thm, no remainder) says:

$$\mathbf{0} = \nabla \log L(\widehat{\theta}) = \nabla \log L(\theta_0) + \nabla\nabla^{tr} \log L(\theta^*)\,(\widehat{\theta} - \theta_0)$$

$$= \nabla \log L(\theta_0) - \Big\{ - \nabla^{\otimes 2} \log L(\theta^*) \Big\} (\widehat{\theta} - \theta_0)$$

**Observed Information:** put $J = -\nabla^{\otimes 2} \log L(\widehat{\theta})$

# Observed Information, II

**Observed Information:** via Law of Large Numbers

$$J \;=\; -\nabla^{\otimes 2} \log L(\widehat{\theta}) \;\approx\; n \cdot E(-\nabla^{\otimes 2} \log f(X_1, \theta_0)) \qquad \text{Fisher Info}$$

So from the previous slide

$$\mathbf{0} \;\approx\; \frac{1}{\sqrt{n}} \nabla \log L(\theta_0) \;-\; \left\{\frac{1}{n} J\right\} \{\sqrt{n}\,(\widehat{\theta} - \theta_0)\} \qquad \text{and}$$

$$\boxed{\sqrt{n}\,(\widehat{\theta} - \theta_0) \;\approx\; \left(n\, J^{-1}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \log f(X_i, \theta_0)}$$

**Observed Information is the Hessian (matrix of 2nd partial deriv's) of negative loglikelihood given by MLE software.**

# Logistic Regression Estimating Equation

**Logistic Regression** is a model for binary $Y_i$ given $X_i$ :

$$\text{logit}\, P(Y_i = 1 \,|\, X_i) \;=\; X_i' \beta \;, \qquad \text{logit}(x) = \log\left(\frac{x}{1-x}\right) \quad \text{log-odds}$$

$$\boxed{P(Y_i = 1 \,|\, X_i,\, \beta) \;=\; e^{\beta' X_i} / (1 + e^{\beta' X_i})} \qquad \texttt{plogis}(x) = \frac{e^x}{1+e^x}$$

**Data** $\{(Y_i, X_i)\}_{i=1}^n$, $\qquad L(\beta) \;=\; \prod_{i=1}^n \left[\left(\frac{e^{\beta' X_i}}{1+e^{\beta' X_i}}\right)^{Y_i} \left(\frac{1}{1+e^{\beta' X_i}}\right)^{1-Y_i}\right]$

**logLik** $\qquad \log L(\beta) \;=\; \sum_{i=1}^n \left[Y_i\, \beta' X_i \;-\; \log\left(1 + e^{\beta' X_i}\right)\right]$

**Equation:** $\qquad \boxed{\nabla \log L(\beta) \;=\; \sum_{i=1}^n \; X_i \left[Y_i - \frac{e^{\beta' X_i}}{1+e^{\beta' X_i}}\right] \;=\; \mathbf{0}}$

Compare least-squares estimating equation !

# Poisson Regression Estimating Equation

**Poisson Regression:** model for Poisson counts $Y_i$ given $X_i$ :

$\log\left\{E(Y_i\,|\,X_i)\right\} = X_i'\beta$ , **Poisson rate** $\lambda_i = e^{\beta' X_i}$ for $Y_i$

$$\boxed{P(Y_i = k\,|\,X_i,\,\beta) \;=\; e^{-\lambda_i}\,\lambda_i^k/\,k! \;=\; \texttt{dpois}(k,\lambda_i)}$$

**Data** $\{(Y_i, X_i)\}_{i=1}^n$, $L(\beta) = \prod_{i=1}^n \left[\exp\left(-\,e^{\beta' X_i}\right) e^{Y_i\,\beta' X_i}\right]$

**logLik** $\log L(\beta) \;=\; \sum_{i=1}^n \left[Y_i\,\beta' X_i \;-\; e^{\beta' X_i}\right]$

**Equation:** $$\boxed{\nabla \log L(\beta) \;=\; \sum_{i=1}^n X_i\left[Y_i \;-\; e^{\beta' X_i}\right] \;=\; \mathbf{0}}$$

# Exponential Families and GLM's

Exponential families have densities $f(y,\theta) = e^{\theta' T(y) - c(\theta)}\, h(y)$

$\theta$ is the **natural parameter** (not always the simplest parameter)

**Examples:** (1) Binom($k, \pi$) , $p(y,\pi) = \binom{k}{y} \pi^y (1-\pi)^{k-y}$

So $p(y,\pi) \propto \exp\Big(y \log(\frac{\pi}{1-\pi}) + k \log(1-\pi)\Big)$ , $\theta = \log(\frac{\pi}{1-\pi})$

and $1 - \pi = (1+e^\theta)^{-1} \;\Rightarrow\; c(\theta) = k \log(1+e^\theta))$

(2) Poisson($\lambda$) , $p(y,\pi) = \frac{\lambda^y}{y!} e^{-\lambda} = \frac{1}{y!} \exp\Big(y \log(\lambda) - \lambda\Big)$

So $\theta = \log(\lambda)$ , $c(\theta) = e^\theta$ in this example.

# Quick Facts about (Natural) Exponential Families

(a) $\frac{d}{d\theta}\sum_y p(y,\theta) = 0 \quad \Rightarrow \quad c'(\theta) = E_\theta(T(Y_1))$

(b) for $Y_1,\ldots,Y_n$, $\quad \log L(\theta) = \sum_{i=1}^n \left[\theta' T(Y_i) - c(\theta)\right]$

is strictly concave with MLE defined uniquely, if it exists, by

$c'(\theta) = E_\theta(T(Y_1)) = n^{-1}\sum_{i=1}^n T(Y_i)$ **Score Eq'n, GMOM est.**

Next step is to combine the modeling ideas from the Logistic and Poisson Regression slides into a unified "Generalized Linear Modeling" framework, introduced in Sec. 4.1 and told in more detail in Sec. 4.4 of Agresti.

# Ingredients and Terminology for GLMs

$Y_i$ response variables satisfying exp. family model $Y_i \sim f(y, \theta_i)$

$X_i$ (vector) regressor variables entering model via $\eta_i = \beta' X_i$

$\mu_i$ cond. expectation of $Y_i$ given $X_i$

$\theta_i$ monotonically related to $\mu_i = \int y f(y, \theta_i) dy$ through model

$g(\mu_i) = \eta_i$ **link function** $g$ monotonic, smooth

GLM contains relationships $\beta \mapsto \eta_i \mapsto \mu_i \mapsto \theta_i$

specifying likelihood $L(\beta) = \prod_{i=1}^{n} f(Y_i, \theta_i)$

## Estimating Equation for GLM

Next time derive in detail the Score Equation $\nabla_\beta \log L(\beta) = \mathbf{0}$

We find that this is an **Estimating Equation** sum of iid terms, sketch theory to show that estimator $\tilde{\beta}$ solving it makes

$$\sqrt{n}(\tilde{\beta} - \beta) \overset{\mathcal{D}}{\to} \mathcal{N}(0, V) \quad \text{as} \quad n \to \infty$$

**if** $g(\mu_i) \equiv \eta_i$ hold for iid data even without the density model.

Some simplifications occur when **link is canonical**, i.e. $\eta_i$ is the natural parameter for the exponential-family (as in our 2 examples).

Now look at R coding, in `glmRcode.RLog`