BIOSTAT 651

Notes #4: Generalized Linear Models

- Topics:
  - Introduction to GLM
  - Exponential families
- Text (Dobson & Barnett, 3rd Ed.): Chapter 3

From Linear Regression to GLM

- Linear regression model:

$$\begin{aligned} Y_i &= \mathbf{x}_i^T \boldsymbol{\beta} + e_i \\ E[Y_i | \mathbf{x}_i] &= \mathbf{x}_i^T \boldsymbol{\beta} \\ V(Y_i | \mathbf{x}_i) &= \sigma^2 \\ Y_i &\sim \text{Normal} \end{aligned}$$

- The *generalization* part of GLM refers to:
  - dropping the Normality requirement
  - relaxing the constant variance assumption
  - allowing for some function of $E[Y_i]$ to be linear in the parameters
- In GLM, the focus is on the *exponential family*
  - members include: Exponential, Poisson, Binomial, Gamma, Normal

Exponential Family

Exponential Family

- If a distribution is an exponential family, then its probability/density function can be written as:

$$f(Y;\theta,\phi) \quad = \quad \exp\left\{\frac{t(Y)\theta - b(\theta)}{a(\phi)} + c(Y,\phi)\right\}$$

  - typically, θ is the parameter of interest

    relates to the mean function
  - in contrast, ϕ (dispersion) is treated as a nuisance parameter related to the variance

- In GLM, we attempt to separate the mean and variance components

- If $t(Y) = Y$, the family is in *canonical form*, in which case θ is referred to as the canonical (*natural*) parameter

Exponential Family (continued)

- Note:
  - for now, we have one θ indexing any Y
  - in the regression setting (later), we replace θ with θ_i

Exponential Family: Binomial

- Suppose $Y \sim \text{Binomial}(n, \pi)$

$$p(Y;\pi) \quad = \quad \begin{pmatrix} n \\ Y \end{pmatrix} \pi^Y (1-\pi)^{n-Y}$$

$$= \exp\left\{ Y \log\left(\frac{\pi}{1-\pi}\right) + n\log(1-\pi) + \log \begin{pmatrix} n \\ Y \end{pmatrix} \right\}$$

- Therefore,

$$t(Y) \quad =$$

$$a(\phi) \quad =$$

$$\theta \quad =$$

$$b(\theta) \quad =$$

$$c(Y,\phi) \quad =$$

Exponential Family: Poisson Case

- Suppose $Y \sim \text{Poisson}(\lambda)$,

$$\begin{aligned} p(Y;\lambda) &= \frac{e^{-\lambda}\lambda^{Y}}{Y!} \\ &= \exp\{Y\log(\lambda) - \lambda - \log(Y!)\} \end{aligned}$$

- Therefore,

$$t(Y) =$$

$$a(\phi) =$$

$$\theta =$$

$$b(\theta) =$$

$$c(Y,\phi) =$$

Exponential Family: Normal

- $Y \sim \text{Normal}(\mu, \sigma^2)$, with σ^2 known

$$
\begin{aligned}
f(Y) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(Y-\mu)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{(Y-\mu)^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})\right\} \\
&= \exp\left\{\frac{2\mu Y - \mu^2 - Y^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})\right\} \\
&= \exp\left\{\frac{\mu Y - (1/2)\mu^2}{\sigma^2} - \frac{Y^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})\right\}
\end{aligned}
$$

such that

$$
\begin{aligned}
t(Y) &= \\
\theta &= \\
a(\phi) &= \\
b(\theta) &=
\end{aligned}
$$

Exponential Family: Likelihood

- For a single data point

$$\begin{aligned} L_i(\theta) &\propto f(Y_i; \theta, \phi) \\ \ell_i(\theta) &= \log L_i(\theta) \end{aligned}$$

- Referring to the previous set-up (canonical form),

$$\ell_i(\theta) = \frac{Y_i\theta - b(\theta)}{a(\phi)}$$

  taking derivatives w.r.t θ,

$$\begin{aligned} U_i(\theta) &= \frac{\partial \ell_i}{\partial \theta} = \frac{Y_i - b'(\theta)}{a(\phi)} \\ J_i(\theta) &= \frac{-\partial^2 \ell_i}{\partial \theta^2} = \frac{b''(\theta)}{a(\phi)} \\ I_i(\theta) &= E[J_i(\theta)] = \frac{b''(\theta)}{a(\phi)} \end{aligned}$$

Exponential Family: Likelihood (continued)

- Properties of the likelihood function:

$$\begin{aligned} E[U_i(\theta)] &= 0 \\ V[U_i(\theta)] &= I_i(\theta) \end{aligned}$$

- Combining these results,

$$E[Y_i] \equiv \mu = b'(\theta)$$

and, in addition,

$$\begin{aligned} \frac{b''(\theta)}{a(\phi)} &= \frac{V(Y_i)}{a(\phi)^2} \\ V(Y_i) &= b''(\theta)a(\phi) \end{aligned}$$

Mean and Variance Functions

- Note that $E[Y_i]$ depends only on the natural parameter, θ
  - although $V(Y_i)$ is a function of both θ and ϕ
- The variance is often expressed as

$$V(Y_i) = v(\mu)a(\phi)$$

  where $v(\mu)$ is written in terms of only μ

- Since we have already derived $V(Y_i) = b''(\theta)a(\phi)$, it follows that

$$v(\mu) = b''(\theta)$$

Exponential Family: Mean and Variance

- e.g., Applying these ideas to the binomial case:

$$
\begin{aligned}
b(\theta) &= n \log(1 + e^{\theta}) \\
b'(\theta) &= n \, \frac{e^{\theta}}{(1 + e^{\theta})} \\
b''(\theta) &= n \, \frac{e^{\theta}}{(1 + e^{\theta})^2}
\end{aligned}
$$

such that

$$
\begin{aligned}
E[Y] &= b'(\theta) = n \, \frac{e^{\theta}}{(1 + e^{\theta})} \\
V(Y) &= b''(\theta) a(\phi) = n \, \frac{e^{\theta}}{(1 + e^{\theta})^2}
\end{aligned}
$$

Mean and Variance (continued)

- e.g., applying to the Normal case:

$$
\begin{aligned}
b(\theta) &= \frac{\theta^2}{2} \\
b'(\theta) &= \theta \\
b''(\theta) &= 1
\end{aligned}
$$

such that

$$
\begin{aligned}
E[Y] &= \theta \\
V(Y) &= \sigma^2
\end{aligned}
$$

General k-Parameter Exponential Family

- Set $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$
- A distribution is a k-parameter exponential family if its probability/density function can be expressed in the following form:

$$f(Y; \boldsymbol{\theta}) = \exp\left\{ \sum_{j=1}^{k} t_j(Y)\theta_j - b(\boldsymbol{\theta}) + c(Y) \right\}$$

- In this setting, all k parameters are of interest

- e.g., Normal (σ^2 unknown)

Regression Modeling Using GLM

Generalized Linear Models

- Initially developed by Nelder & Wedderburn (1972, *JRSSA*)
  - assume that a known function of $\mu_i = E[Y_i]$ is related linearly to $\mathbf{x}_i$

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

  - $g(\cdot)$ is referred to as the *link* function
- Still assume independence of $Y_1, \ldots, Y_n$
- Linearity assumption now applies to $g(\mu_i)$, which need not equal $E[Y_i]$

Components of the GLM

- In setting up a GLM, the following are specified:

1. Distribution (random component)
   - Y_i assumed to follow a (canonical) exponential family:

$$f(Y_i; \theta_i, \phi) = \exp\left\{ \frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i, \phi) \right\}$$

2. Systematic component
   - linear predictor: $\eta_i \equiv \mathbf{x}_i^T \boldsymbol{\beta}$

3. Link function
   - connects $\mathbf{x}_i$ and μ_i
   - $g(\mu_i) = \eta_i$
   - required that g be monotone, differentiable function

     $g^{-1}(\eta_i) = \mu_i$

Link Functions

- Commonly chosen link functions include

$$\begin{array}{rl}
\text{log} & \eta_i = \log(\mu_i) \\[2em]
\text{logit} & \eta_i = \log\left\{\dfrac{\mu_i}{1-\mu_i}\right\} \\[2em]
\text{probit} & \eta_i = \Phi^{-1}(\mu_i) \\[2em]
\text{complementary} & \\
\text{log-log} & \eta_i = \log\{-\log(1-\mu_i)\}
\end{array}$$

where $\Phi(\cdot)$ is the CDF for a N(0,1) variate

Canonical Link

- We observe $(Y_i, \mathbf{x}_i)$ for $i = 1, \ldots, n$, where the distribution of $(Y_i|\mathbf{x}_i)$ is assumed to be of the form

$$f(Y_i; \theta_i, \phi) \quad = \quad \exp\left\{\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i, \phi)\right\}$$

- Using previously described properties of exponential families:

$$\begin{aligned} E[Y_i] = \mu_i &= b'(\theta_i) \\ V(Y_i) &= b''(\theta_i)a(\phi) = v(\mu_i)a(\phi) \end{aligned}$$

- Link function, $g(\cdot)$, is canonical if $\eta_i = \theta_i$

- Note: the canonical link is usually preferred due to some desirable statistical and computational properties.

Range Restrictions

- In linear regression, $\mu_i \in (-\infty, \infty)$ and $\mathbf{x}_i^T \boldsymbol{\beta} \in (-\infty, \infty)$
  - in fact, $g(\mu_i) = \mu_i$ (identity link) is typically chosen when $Y_i \sim$ Normal
- For links other than the identity, range restrictions should be accommodated
  - e.g., for $Y_i \sim$ Poisson, $\mu_i > 0$

    select $\mu_i = e^{\eta_i} > 0$
  - e.g., for $Y_i \sim$ Bernoulli, $\mu_i \in (0, 1)$

    select $\mu_i = e^{\eta_i}/\{1 + e^{\eta_i}\} \in (0, 1)$
  - in both cases, canonical link

Deriving Canonical Link

- Examples: deriving the canonical link:
  - e.g., $Y_i \sim$ Normal
  - e.g., $Y_i \sim$ Bernoulli
  - e.g., $Y_i \sim$ Poisson

Choice of Link Function

- It is possible to use links that are not canonical
- e.g., possible that $Y_i \sim$ Normal, but that covariate effects are multiplicative
  - implies $\mu_i = e^{\eta_i}$
- e.g., $Y_i \sim$ Poisson, but with additive covariate effects
  - implies $\mu_i = \eta_i$
  - preferably, $\widehat{\mu}_i < 0$ never, or rarely
- Some would argue that the link function should be chosen in accordance with the investigator's objectives