B.4 Generalized Linear Models

B.4.1 Linear Exponential Family
GLM Intro
Linear Exponential Family
Exercises

Sources and Objectives

This lesson comes from Frees sections 13.1-13.2.1 and is designed to address SRM Syllabus learning objective 2.

GLM Ingredients

Goal: Single framework for all regression models

1. *Systematic or linear component*: $\beta_0 + \beta_1 X_i$, (sometimes called η_i)
2. *Link function* $g(\mathrm{E}[Y]) = \beta_0 + \beta_1 X_i$
   - Normal response: $\mu = \beta_0 + \beta_1 X_i \longrightarrow g(\mu) = \mu$, the *identity link*
   - Binary response: $g(\mu) = \ln(\mu/(1-\mu)) = \beta_0 + \beta_1 X_i$, the *logit link*
   - Poisson response: $g(\mu) = \ln(\mu) = \beta_0 + \beta_1 X_i$, the *log link*
3. *Response distribution or random component*.

   So far, $N(\mu, \sigma^2)$, Binary, Multinomial, or Count Response

   $\longrightarrow$ any linear exponential family member

As before, we assume

- Explanatory variables are not random
- Observations should be independent

Normal Distribution

The Score equations for the Normal, Binomial, and Poisson cases all turned out to be

$$\sum(y_i - \hat{\mu}) = 0 \quad \sum(y_i - \hat{\mu})x_i = 0$$

These come from log-likelihoods, so let's find the common thread.

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

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

Collect terms involving μ - these don't disappear in ℓ'.

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

Poisson and Binary Distribution

Count Data: Poisson Distribution

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

Binary Response: Bernoulli Distribution:

$$f(y;\pi) = \pi^y(1-\pi)^{1-y}$$
$$= \exp\left(y\ln(\pi) - y\ln(1-\pi) + \ln(1-\pi)\right)$$
$$= \exp\left([y\ln(\pi/(1-\pi)) + \ln(1-\pi)] + 0\right)$$

What's the pattern?

$$f_N(y;\mu,\sigma^2) = \exp\left(\left[\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2}\right] - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right)$$
$$f_P(y;\mu) = \exp\left([y\ln(\mu) - \mu] - \ln(y!)\right)$$
$$f_B(y;\pi) = \exp\left([y\ln(\pi/(1-\pi)) + \ln(1-\pi)] + 0\right)$$
$$f(y;\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + S(y,\phi)\right)$$

Natural or canonical parameter θ locates the distribution

$$\theta_N = \mu, \qquad \theta_P = \ln(\mu), \qquad \theta_B = \ln(\pi/(1-\pi))$$

Dispersion parameter ϕ *or* ϕ/w scales the distribution

$$\phi_N = \sigma^2, \qquad \phi_P = 1, \qquad \phi_B = 1$$

Cumulant function $b(\theta)$ specifies mean, variance

$$b_N(\theta) = \tfrac{1}{2}\theta^2, \qquad b_P(\theta) = e^\theta, \qquad b_B(\theta) = \ln(1+e^\theta)$$

Leftover terms $S(y,\phi)$ disappears when we compute ℓ'.

Mean

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

Mean:

$$\mathrm{E}\left[\tfrac{d}{d\theta}\ell(y;\theta)\right] = \int \frac{f'(y;\theta)}{f(y;\theta)} \cdot f(y;\theta)\, dy \overset{E.F.}{=} \frac{d}{d\theta}\int f(y;\theta)\, dy = 0.$$

$$\ell(y;\theta) = (y\theta - b(\theta))/\phi + S(y,\phi) \longrightarrow \ell'(y;\theta) = (y - b'(\theta))/\phi$$

$$\mathrm{E}\left[(Y - b'(\theta))/\phi\right] = 0 \longrightarrow \mathrm{E}[Y] = b'(\theta)$$

Test against examples:

$$b'_N(\theta) = \theta\checkmark \qquad b'_P(\theta) = e^\theta = \mu\checkmark \qquad b'_B(\theta) = \tfrac{e^\theta}{1+e^\theta} = \pi\checkmark$$

Variance

Variance: By similar likelihood trick:

$$\mathrm{Var}(Y) = b''(\theta)\cdot\phi \text{ (or } \phi/w \text{ if weights are used)}$$

Test against examples:

$$b''_N(\theta)\cdot\phi = 1\cdot\sigma^2\checkmark$$

$$b''_P(\theta)\cdot\phi = e^\theta\cdot 1 = \mu\checkmark$$

$$b''_B(\theta)\cdot\phi = \frac{e^\theta}{(1+e^\theta)^2}\cdot 1 = \pi(1-\pi)\checkmark$$

Exercise 1

For a gamma(α, γ) distribution, $f(y) = \dfrac{(y\gamma)^\alpha e^{-y\gamma}}{y\Gamma(\alpha)}$. Put this density in exponential form. If $\phi = 1/\alpha$, what is the natural parameter?

Exercise 1

For a gamma(α, γ) distribution, $f(y) = \dfrac{(y\gamma)^\alpha e^{-y\gamma}}{y\Gamma(\alpha)}$. Put this density in exponential form. If $\phi = 1/\alpha$, what is the natural parameter?

$$
\begin{aligned}
f(y) &= e^{\alpha \ln(y\gamma) - y\gamma + \text{stuff}} \\
&= e^{y(-\gamma) + \alpha \ln(\gamma) + \alpha \ln(y) + \text{stuff}} \\
&= e^{\dfrac{y(-\gamma/\alpha) - (-\ln(\gamma))}{1/\alpha} + \text{stuff}} \longrightarrow \theta = \boxed{-\frac{\gamma}{\alpha}} = -\frac{1}{\mu}.
\end{aligned}
$$

What about $b(\theta)$?

$\ln(\gamma) = \ln(\alpha\gamma/\alpha) = \ln(-\theta) + \ln(\alpha)$

- roll $\ln(\alpha)$ bit in with the "stuff" to make $S(y, \phi)$
- Then $b(\theta) = -\ln(-\theta)$

Exercise 2

You are given:

- A response variable Y is in the exponential family,

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

- $b(\theta) = e^{\theta}$
- $\mu = \mathrm{E}[Y]$
- Dispersion parameter $\phi = 1$

Determine an expression for $\mathrm{Var}(Y)$ as a function of μ.

A. μ^2 B. μ C. $1/\mu$ D. 1 E. e^{μ}

Exercise 2

You are given:

- A response variable Y is in the exponential family,

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

- $b(\theta) = e^{\theta}$
- $\mu = \mathrm{E}[Y]$
- Dispersion parameter $\phi = 1$

Determine an expression for $\mathrm{Var}(Y)$ as a function of μ.

A. μ^2 B. μ C. $1/\mu$ D. 1 E. e^{μ}

$\mathrm{Var}(Y) = \phi b''(\theta) = 1 \cdot e^{\theta}$, but that's not in terms of μ.

$\mu = \mathrm{E}[Y] = b'(\theta) = e^{\theta} \qquad \longrightarrow \mathrm{Var}(Y) = 1 \cdot e^{\theta} = \boxed{\mu}$