Generalized Linear Models (GLMs)

Bruce A Craig

Department of Statistics
Purdue University

Outline

- Exponential family distributions
    - Examples of distributions
    - Expected value and variance
- Generalized linear models
    - Link functions
    - Estimation

Material covered in Chapter 8 of Faraway textbook

LMs versus GLMs

- Linear model
  - Conceptual framework "DATA = FIT + ERROR"
  - FIT: Means of $\mathbf{Y}$ are a linear function of $\mathbf{X}$
  - ERROR: Deviations $\varepsilon \sim N(0, \sigma^2\mathbf{I})$
  - Overall model can be expressed $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$
  - Normality most critical with prediction intervals
- Generalized linear model
  - The $\mathbf{Y}$ are from an exponential family distribution
  - The means of $\mathbf{Y}$ are linked to a linear function of $\mathbf{X}$
  - Variance of each Y often a function of its mean
  - Best to move away from "DATA = FIT + ERROR" conceptual framework

Exponential Family Distribution (EFD)

- Distribution can be discrete or continuous
- The probability mass/density function has general form

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

- θ_i is the *canonical parameter* representing location (also called the *natural parameter*)
- ϕ is the *dispersion parameter* representing scale
- $a(\cdot), b(\cdot), c(\cdot)$ known functions
- Usually, $a(\phi) = \phi/w$
    - w is a known weight, can vary between observations
    - ϕ can be known (one-parameter distribution) or unknown (two-parameter distribution)

Some Exponential Family Dists

- Discrete
  - Binomial (with known N)
  - Poisson
  - Negative binomial (with known r)
  - Multinomial
- Continuous
  - Exponential
  - Normal (and multivariate Normal)
  - Gamma
  - Beta
  - Weibull (with known shape)
  - Inverse gaussian

Poisson Distribution $Y_i \sim Poi(\lambda)$

- The PMF is

$$\begin{aligned} f(y_i) &= \frac{\exp\{-\lambda\}\lambda^{y_i}}{y_i!} \\ &= \exp\left\{y_i \log(\lambda) - \lambda - \log(y_i!)\right\} \end{aligned}$$

- $\theta = \log(\lambda)$, $\phi = 1$, and $w = 1$.
- $a(\phi) = \phi/w$, $b(\theta) = \exp\{\theta\}$, and $c(y_i, \phi) = -\log(y_i!)$

Binomial Distribution $Y_i \sim bin(N, p)$

- The PMF is

$$
\begin{aligned}
f(y_i) &= \binom{N}{y_i} p^{y_i}(1-p)^{N_i - y_i} \\
&= \exp\left\{ y_i \log \frac{p}{1-p} + N\log(1-p) + \log \binom{N}{y_i} \right\}
\end{aligned}
$$

- $\theta = \log \frac{p}{1-p}$, $\phi = 1$, and $w = 1$
- $a(\phi) = \phi / w$, $b(\theta) = Nlog\left(1 + exp(\theta)\right)$, $c(y_i, \phi) = \log \binom{N}{y_i}$

Binomial Distribution $Y_i \sim bin(N, p)$

- The PMF for $\hat{p}_i = y_i/N$ is

$$
\begin{aligned}
f(\hat{p}_i) &= \binom{N}{N\hat{p}_i} p^{N\hat{p}_i}(1-p)^{N(1-\hat{p}_i)} \\
&= \exp\left\{ \frac{\hat{p}_i \log \frac{p}{1-p} + \log(1-p)}{1/N} + \log \binom{N}{N\hat{p}_i} \right\}
\end{aligned}
$$

- $\theta = \log \frac{p}{1-p}$, $\phi = 1$, and $w = N$
- $a(\phi) = \phi/w$, $b(\theta) = \log[1 + exp(\theta)]$, $c(\hat{p}_i, \phi) = \log \binom{N}{N\hat{p}_i}$

Normal Distribution $Y_i \sim N(\mu, \sigma)$

- The PDF is

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

- $\theta = \mu$, $\phi = \sigma^2$, and $w = 1$
- $a(\phi) = \phi/w$, $b(\theta) = \theta^2/2$, $c(y_i, \phi) = -\left(y_i^2/\phi + \log(2\pi\phi)\right)/2$

Gamma Distribution $Y_i \sim \Gamma(\alpha, \beta)$

- The PDF is

$$
\begin{aligned}
f(y_i) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} \exp\{-\beta y_i\} \\
&= \exp\{(\alpha-1)\log y_i - \beta y_i + \alpha \log \beta - \log \Gamma(\alpha)\} \\
&= \exp\left\{ \frac{y_i \frac{\beta}{\alpha} - \log(\beta)}{-1/\alpha} + (\alpha-1)\log y_i - \log \Gamma(\alpha) \right\}
\end{aligned}
$$

- $\theta = \beta/\alpha$, $\phi = 1/\alpha$, and $w = 1$
- $a(\phi) = -\phi/w$, $b(\theta_i) = \log(\theta)$,
  $c(y_i, \phi) = (1/\phi - 1)\log y_i - \log \Gamma(1/\phi) - \log \phi/\phi$

Expected Value of EFD

- Can use likelihood theory to determine mean and variance
- Given $Y \sim EFD(\theta, \phi)$, the log likelihood is

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

- We know $E(l'(\theta)) = 0$ at true value of θ, so

$$\begin{aligned} l'(\theta, \phi; y) &= \frac{y - b'(\theta)}{a(\phi)} \\ &\downarrow \\ 0 &= \frac{E(y) - b'(\theta)}{a(\phi)} \end{aligned}$$

- Thus, $E(Y) = b'(\theta)$

Variance of EFD

- Similarly, we know

$$\begin{aligned} E(l''(\theta)) &= -E[(l'(\theta))^2] \\ &= -E\left(\frac{(y-b'(\theta))^2}{a^2(\phi)}\right) \\ &= -Var(Y)/a^2(\phi) \end{aligned}$$

- On the left-hand side

$$\begin{aligned} l''(\theta,\phi;y) &= \frac{-b''(\theta)}{a(\phi)} \\ &\downarrow \\ E(l''(\theta)) &= \frac{-b''(\theta)}{a(\phi)} \end{aligned}$$

- Thus, $Var(Y) = b''(\theta)a(\phi)$

Back to Examples

- Poisson

$$
\begin{aligned}
E(Y) &= b'(\theta) = \exp\{\log\lambda\} = \lambda \\
Var(Y) &= b''(\theta)a(\phi) = \exp\{\log\lambda\} \times 1 = \lambda
\end{aligned}
$$

- Binomial

$$
\begin{aligned}
E(Y) &= b'(\theta) = N\frac{\exp\{\theta\}}{1+\exp\{\theta\}} = Np \\
Var(Y) &= b''(\theta)a(\phi) = N\frac{\exp\{\theta\}}{(1+\exp\{\theta\})^2} \times 1 = Np(1-p)
\end{aligned}
$$

- Binomial

$$
\begin{aligned}
E(\hat{p}) &= b'(\theta) = \frac{\exp\{\theta\}}{1+\exp\{\theta\}} = p \\
Var(\hat{p}) &= b''(\theta)a(\phi) = \frac{\exp\{\theta\}}{(1+\exp\{\theta\})^2} \times \frac{1}{N} = p(1-p)/N
\end{aligned}
$$

Back to Examples

- Normal

$$\begin{aligned} E(Y) &= b'(\theta) = \theta = \mu \\ Var(Y) &= b''(\theta)a(\phi) = 1 \times \sigma^2 = \sigma^2 \end{aligned}$$

- Gamma

$$\begin{aligned} E(Y) &= b'(\theta) = \frac{1}{\theta} = \alpha/\beta \\ Var(Y) &= b''(\theta)a(\phi) = -\theta^{-2} \times \frac{-1}{\alpha} = \alpha/\beta^2 \end{aligned}$$

Generalized Linear Models

- Used to describe relationship between observations from an **EFD** and a set of predictors **x**
- Includes linear models so this is a broader more general model framework
- In addition to the specific distribution, need to specify a **link function** that describes how the mean of the response is related to a linear combination of predictors
- Do not confuse the link function with a transformation of the response variable
    - Analysis performed on the data, or Y, scale
    - Parameters β, however, are on the link scale

Generalized Linear Models

- Data: $(y_i, \mathbf{x}_i) = (\ y_i, x_{i1}, x_{i2}, \cdots, x_{i,p-1}\ )$, $i = 1, 2, \cdots, n$
- Random component: $y_i \mid \mathbf{x}_i \overset{ind}{\sim} EFD(\theta_i, \phi)$
- Systematic component: Joint effects of $\mathbf{x}_i$ determined through a linear combination

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} = \mathbf{x}_i \beta$$

- Link function $g(\mu_i)$ relates $\mu_i = E\{y_i\}$ with $\eta_i = \mathbf{x}_i \beta$

$$g(\mu_i) = \eta_i$$

Link Functions

- The link function $g(\mu_i)$ completes a probability model
  - Logistic regression: $g(\mu_i) = logit(\mu_i) = \mathbf{x}_i\beta$
  - Probit regression: $g(\mu_i) = \Phi^{-1}(\mu_i) = \mathbf{x}_i\beta$
  - Poisson regression: $g(\mu_i) = \log \mu_i = \mathbf{x}_i\beta$
  - Normal regression: $g(\mu_i) = \mu_i = \mathbf{x}_i\beta$
- As it defines the mean function for the EFD
  - Logistic regression: $\mu_i = \frac{1}{1+\exp\{-\mathbf{x}_i\beta\}} = \frac{\exp\{\mathbf{x}_i\beta\}}{1+\exp\{\mathbf{x}_i\beta\}}$
  - Probit regression: $\mu_i = \Phi(\mathbf{x}_i\beta)$
  - Poisson regression: $\mu_i = \exp\{\mathbf{x}_i\beta\}$
  - Normal regression: $\mu_i = \mathbf{x}_i\beta$

Canonical Link Functions

- Any monotone, continuous, and differentiable function can serve as the link
- When $\eta = g(\mu) = \theta$, then the canonical link is used
- This means that $g(b'(\theta)) = \theta$ and $\mathbf{X}'\mathbf{Y}$ is sufficient for $\boldsymbol{\beta}$

Normal: identity	Poisson: log
Binomial: logit	Gamma: inverse

- While computationally convenient, context can compel using something other than the canonical link

Fitting a GLM

- Estimation based on **maximum likelihood**
- Log-likelihood:

$$
\begin{aligned}
l(\beta) &= \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \\
&\downarrow \\
\frac{\partial l}{\partial \beta_j} &= \frac{1}{\phi} \sum_{i=1}^{n} w_i \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - b'(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} \right)
\end{aligned}
$$

- Using chain rule and fact that $\partial \mu_i / \partial \theta_i = b''(\theta_i)$

$$
\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^{n} \frac{(y_i - b'(\theta_i))}{b''(\theta_i)/w_i} \frac{\partial \mu_i}{\partial \beta_j}
$$

Fitting a GLM, continued

- Using mean and variance relationships, we set these partial derivatives to 0 to get the score equations

$$0 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(\mu_i)} x_{ij}/g'(\mu_i)$$

- Use iterative reweighted least-squares to obtain estimates
  1. Set $k = 1$ and initial guess for $\boldsymbol{\beta}$ $(\boldsymbol{\beta} \rightarrow \hat{\boldsymbol{\eta}} \stackrel{g^{-1}}{\rightarrow} \hat{\boldsymbol{\mu}})$
  2. Approx $g(y_i) = z_i$ using linearity around $\hat{\mu}^{(k)}$
  $$z_i^{(k)} = g(\hat{\mu}_i^{(k)}) + (y_i - \hat{\mu}_i^{(k)})g'(\hat{\mu}_i^{(k)})$$
  3. Form weights using $\text{Var}(z_i^{(k)}) = (g'(\hat{\mu}^{(k)}))^2 \, \text{V}(\hat{\mu}_i^{(k)})$
  4. Estimate $\hat{\beta}^{(k+1)} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Z}$
  5. Repeat Steps 2-4 until convergence

Maximum Likelihood

- Likelihood function
  - describes how probable data are for diff parameter values
- ML estimates
  - The parameter values that maximize the likelihood
  - Under regularity conditions, $cov(\hat{\boldsymbol{\beta}})$ is the inverse of the information matrix

$$\left[-E\left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_i\partial\beta_j}\right)\right]^{-1}$$

- Properties
  - large-sample Normal distributions
  - asymptotically consistent (i.e. converge to the true parameter as sample size increases)
  - asymptotically efficient (i.e. large-sample SE no greater than other estimation methods)

Fitting a GLM in R

- Utilize the `glm` and related functions

```
glm(formula, family = gaussian, data,...)
```

- Use `family` option to define EFD and link

```
family = binomial(link = "logit")
family = gaussian(link = "identity")
family = gamma(link = "inverse")
family = inverse.gaussian(link = "1/mu^2")
family = poisson(link = "log")
```