STAT347: Generalized Linear Models
Lecture 2

Today's topics: Agresti Chapters 4.1-4.2

- The exponential dispersion family
- Likelihood score equations for parameter estimation

1 The exponential dispersion family

1.1 Definition

A random variable Y follows an exponential dispersion family distribution and has the density $f(y;\theta,\phi)$ of the form ("density" here including the possibility of discrete atoms.)

$$f(y;\theta,\phi) = e^{\frac{y\theta - b(\theta)}{a(\phi)}} f_0(y;\phi)$$

Terminologies:

- θ: natural or canonical parameters
- $b(\theta)$: normalizing or cumulant function
- ϕ: dispersion parameter with $a(\phi) > 0$
- Typically $a(\phi) \equiv 1$ and $f_0(y;\phi) = f_0(y)$. An exception is the Gaussian distribution where $a(\phi) = \sigma^2$

1.2 Some well-known one-parameter exponential families

1. Normal with mean μ and variance σ^2:

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

2. Bernoulli with probability p:

$$\begin{aligned} f(y;p) = p^y(1-p)^{1-y} &= e^{y\log\frac{p}{1-p} + \log(1-p)} \\ &= e^{y\theta - \log[1+e^\theta]} \end{aligned}$$

3. Binomial with p and n:

$$\begin{aligned} f(y;p,n) = \binom{n}{y} p^y (1-p)^{n-y} &= e^{y\log\frac{p}{1-p} + n\log(1-p)} \binom{n}{y} \\ &= e^{y\theta - n\log[1+e^\theta]} \binom{n}{y} \end{aligned}$$

4. Poisson with mean λ:

$$f(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} = e^{y\log\lambda-\lambda}\frac{1}{y!} = e^{y\theta-e^{\theta}}\frac{1}{y!}$$

1.3 Moment relationships

Take the first and second derivative respect to θ_i for both sides of the equation

$$e^{b(\theta)/a(\phi)} = \int e^{y\theta/a(\phi)} f_0(y;\phi)dy$$

We can derive:

$$\mu = \mathbb{E}(Y) = b'(\theta)$$
$$V_\theta = \mathrm{Var}(Y) = b''(\theta)a(\phi)$$

In addition, this indicates that:

$$\frac{\partial\mu}{\partial\theta} = \frac{\mathrm{Var}(Y)}{a(\phi)} > 0$$

thus the mapping from θ to μ is one to one increasing.

2 Exponential family distribution for GLM

- Assume that each observation y_i follows an exponential family with the canonical parameter θ_i and a shared dispersion parameter ϕ
- The mean μ_i of each observation y_i is a function of X_i, so θ_i is also a function of X_i
- The canonical link function in GLM:

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

  As $\mu_i = b'(\theta_i)$, the link function has the form

  $$g(\cdot) = (b')^{-1}(\cdot)$$

  which is called the canonical link.

Canonical link functions for Binomial, Poisson and Bernoulli distributions (will be discussed in lecture).

3 Likelihood score equations

Assume each observation y_i follows an exponential dispersion distribution

$$f(y_i;\theta_i,\phi) = e^{\frac{y_i\theta_i-b(\theta_i)}{a(\phi)}} f_0(y_i;\phi)$$

and the link function $g(\mu_i) = X_i^T\beta$. Then for n independent observations, the log likelihood is

$$L = \sum_i L_i = \sum_i \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_i \log f_0(y_i;\phi)$$

3.1 For the canonical link

If $g(\mu_i) = \theta_i = X_i^T\beta$, then

$$L = \frac{1}{a(\phi)}\left[\sum_j(\sum_i y_i x_{ij})\beta_j - \sum_i b(X_i^T\beta)\right] + \sum_i \log f_0(y_i;\phi)$$

- Score equation for β_j

  $$\frac{\partial L}{\partial \beta_j} = \frac{1}{a(\phi)}\left[\sum_i y_i x_{ij} - \sum_i b'(X_i^T\beta)x_{ij}\right] = \frac{1}{a(\phi)}\left[\sum_i (y_i - \mu_i)x_{ij}\right] = 0$$

  which is equivalent to

  $$\sum_i (y_i - \mu_i)x_{ij} = 0$$

- score equation for a Poisson and Gaussian canonical link model (Section 4.2.2)
  Gaussian model:
  $$\sum_i (y_i - X_i^T\beta)x_{ij} = 0$$
  Poisson model:
  $$\sum_i (y_i - e^{X_i^T\beta})x_{ij} = 0$$

- L is a concave function of β:

  $$\frac{\partial}{\partial \beta}\left[\sum_i (y_i - \mu_i)X_i\right] = -\sum_i \frac{\partial \mu_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta}X_i^T = -\sum_i \frac{\mathrm{Var}(y_i)}{a(\phi)}X_iX_i^T \prec 0$$

3.2 For a general link

Let $\eta_i = g(\mu_i) = X_i^T\beta$ Then

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}$$

We have

- $\frac{\partial L_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)}$
- $\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} = \frac{a(\phi)}{\mathrm{Var}(y_i)}$
- $\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial g(\mu_i)} = \frac{1}{g'(\mu_i)}$
- $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$

Thus, the score equation

$$\frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)}\frac{1}{g'(\mu_i)} = 0$$

- The score equation only depends on the mean and variance of y

- Matrix form of the score equation:

$$\dot{L}(\beta) = X^T D V^{-1}(y - \mu) = 0$$

  where $V = \mathrm{diag}(\mathrm{Var}(y_1), \cdots, \mathrm{Var}(y_n))$ and $D = \mathrm{diag}(g'(\mu_1), \cdots, g'(\mu_n))^{-1}$, $y = (y_1, \cdots, y_n)$ and $\mu = (\mu_1, \cdots, \mu_n)$.

- L is not necessarily a concave function of β.

- A special case: if g is the canonical link, then $D = \frac{1}{a(\phi)} V$ (as $g'(\mu_i) = \partial\theta_i/\partial\mu_i$), thus the score equation is simplified to

$$\frac{1}{a(\phi)} X^T (y - \mu) = 0$$

  the same as what we derived earlier.

- Another special case: if we assume $g(\mu_i) = \mu_i = X_i^T \beta$, then the estimation equations become

$$\sum_i \frac{(y_i - X_i^T \beta) X_i}{\mathrm{Var}(y_i)} = 0$$

Next time: asymptotic distribution of GLM, Hypothesis testing