MA50260 Statistical Modelling

Lecture 9: Exponential Family & GLM

Ilaria Bussoli

March 5, 2024

Types of Response Variables

We considered the following cases:

- **Continuous** $\rightarrow$ Normal
- **Count (bounded)** $\rightarrow$ Binomial
- **Count (unbounded)** $\rightarrow$ Poisson
- **Binary** $\rightarrow$ Bernoulli
- **Time-to-Event** $\rightarrow$ Exponential, Gamma

Generalized Linear Models

A GLM is generally defined by three components:

- The **linear predictor** $\eta_i = \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p} = \mathbf{x}_i^{\mathrm{T}} \underline{\beta}$. This is known as the **systematic component**.
- $g(\mu_i) = \eta_i$ – a **link function** mapping the linear predictor to the mean of the distribution.
- **Probability distribution** $Y_i \sim F(\mu_i)$ from **the exponential family**. The distribution may also have a parameter ϕ. This is termed the **random component**.

The Exponential Family

We say that Y has a distribution in the exponential family if its density can be written

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

where

- θ and ϕ are the **canonical** and **scale** parameters.
- $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specified functions.
- The range of Y doesn't depend on the parameters.

This is known as the **canonical** form.

Normal Example

For $Y \sim \text{Normal}(\mu, \sigma^2)$:

- $\theta = \mu$
- $\phi = \sigma^2$
- $b(\theta) = \frac{\theta^2}{2} = \frac{\mu^2}{2}$
- $a(\phi) = \phi = \sigma^2$
- $c(y, \phi) = -\frac{y^2}{2\phi} - \log \sqrt{2\pi\phi}$

Interpretation of $a(\cdot)$ and $b(\cdot)$ (I)

Consider $\ell(\theta \mid Y) = \log f(Y \mid \theta, \phi)$. Then,

$$\mathbb{E}_\theta\left(\frac{\partial \ell(\theta \mid Y)}{\partial \theta}\right) = 0$$

and

$$\mathrm{Var}_\theta\left(\frac{\partial \ell(\theta \mid Y)}{\partial \theta}\right) = \mathbb{E}_\theta\left(-\frac{\partial^2 \ell(\theta \mid Y)}{\partial \theta^2}\right).$$

If Y has a distribution in the exponential family,

$$\frac{\partial \ell(\theta \mid Y)}{\partial \theta} = \frac{\partial}{\partial \theta}\left\{\frac{Y\theta - b(\theta)}{a(\phi)} + c(Y, \phi)\right\} = \frac{Y - b'(\theta)}{a(\phi)}$$

and

$$\frac{\partial^2 \ell(\theta | Y)}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}.$$

Interpretation of $a(\cdot)$ and $b(\cdot)$ (II)

Taking expectations and solving the equations gives

$$\mathbb{E}(Y) = b'(\theta)$$

and

$$\mathrm{Var}(Y) = a(\phi)b''(\theta).$$

We can write $\mathrm{Var}(Y) = a(\phi)V(\mu)$, with $V(\mu)$ being called the **variance function**.

For the examples in this course, $a(\phi) = \phi/w$, i.e., $\mathrm{Var}(Y) = \phi b''(\theta)/w$.

The function of μ_i obtained by inverting $\mu_i = \mathbb{E}(Y_i) = b'(\theta_i)$ is a commonly-used choice of g. Such a g is called the **canonical link function**.

Estimation of $\underline{\beta}$ (I)

How can we estimate the regression coefficients $\beta_1, \ldots, \beta_p$?

For linear models, we considered the sum of squares

$$S(\underline{\beta}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{i,j} \right)^2$$

to estimate the regression coefficients $\underline{\beta}$.

Estimation of $\underline{\beta}$ (I)

How can we estimate the regression coefficients $\beta_1, \ldots, \beta_p$?

For linear models, we considered the sum of squares

$$S(\underline{\beta}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{i,j} \right)^2$$

to estimate the regression coefficients $\underline{\beta}$.

However: In a GLM framework, y_i and η_i are usually on different scales.

We will use maximum likelihood to estimate the GLM parameters.

For all practical purposes, we assume that $a_i(\phi) = \phi / w_i$ in the exponential family form.

Estimation of $\underline{\beta}$ (II)

Due to independence between observations, the log-likelihood is

$$\ell(\underline{\theta}, \phi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \log f(y_i \mid \theta_i, \phi).$$

Estimation of $\underline{\beta}$ (II)

Due to independence between observations, the log-likelihood is

$$\ell(\underline{\theta}, \phi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \log f(y_i \mid \theta_i, \phi).$$

Using the exponential family form of the density, we write this as

$$\ell(\underline{\theta}, \phi \mid y_1, \ldots, y_n) = \frac{1}{\phi} \left\{ \sum_{i=1}^{n} w_i [y_i \theta_i - b(\theta_i)] \right\} + \sum_{i=1}^{n} c(y_i, \phi).$$

Estimation of $\underline{\beta}$ (II)

Due to independence between observations, the log-likelihood is

$$\ell(\underline{\theta}, \phi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \log f(y_i \mid \theta_i, \phi).$$

Using the exponential family form of the density, we write this as

$$\ell(\underline{\theta}, \phi \mid y_1, \ldots, y_n) = \frac{1}{\phi} \left\{ \sum_{i=1}^{n} w_i [y_i \theta_i - b(\theta_i)] \right\} + \sum_{i=1}^{n} c(y_i, \phi).$$

We will use numerical optimization to find the MLE.