Lecture 3: Exponential Family and GLM

Maochao Xu

Department of Mathematics
Illinois State University
mxu2@ilstu.edu

Exponential family

Consider a single random variable Y whose probability distribution depends on a single parameter θ. The distribution belongs to the exponential family if it can be written in the form

$$f(y;\theta) = s(y)t(\theta)\exp(a(y)b(\theta)),$$

where a, b, s and t are known functions. Equivalently,

$$f(y;\theta) = \exp(a(y)b(\theta) + c(\theta) + d(y)),$$

where $c(\theta) = log(t(\theta))$ and $d(y) = log(s(y))$.

- Canonical (i.e. standard) form:

$$a(y) = y.$$

- Natural parameter of the distribution:

$$b(\theta).$$

Exponential family

Many well-known distributions belong to the exponential family. For example, the Poisson, Normal and Binomial distributions can all be written in the canonical form.

Distribution	Natural parameter	c	d
Poisson	$\log\theta$	$-\theta$	$-\log y!$
Normal	$\dfrac{\mu}{\sigma^2}$	$-\dfrac{\mu^2}{2\sigma^2} - \dfrac{1}{2}\log\left(2\pi\sigma^2\right)$	$-\dfrac{y^2}{2\sigma^2}$
Binomial	$\log\left(\dfrac{\pi}{1-\pi}\right)$	$n\log(1-\pi)$	$\log\binom{n}{y}$

Properties of distributions in the exponential family

- Mean

$$E(a(Y)) = -\frac{c'(\theta)}{b'(\theta)}.$$

- Variance

$$Var(a(Y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}.$$

- Score statistic

$$U = a(Y)b'(\theta) + c'(\theta).$$

- Information

$$\mathcal{J} = Var(U) = \frac{b''(\theta)c'(\theta)}{b'(\theta)} - c''(\theta).$$

or

$$\mathcal{J} = Var(U) = -E(U').$$

Generalized linear model

This model is defined in terms of a set of independent random variables $Y_1, \ldots, Y_N$, each with a distribution from the exponential family and the following properties:

- The distribution of each Y_i has the canonical form and depends on a single parameter θ_i (the θ_i's do not all have to be the same); thus,

$$f(y_i; \theta_i) = \exp(y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)).$$

- The distributions of all the Y_i's are of the same form (e.g., all Normal or all Binomial) so that the subscripts on b, c and d are not needed.
- Suppose that $E(Y_i) = \mu_i$, where μ_i is some function of θ_i. For a generalized linear model there is a transformation of μ_i such that

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

  - Link function: g is a monotone, differentiable function.
  - and

The vector $\mathbf{x}_i$ is a $p \times 1$ vector of explanatory variables (covariates and dummy variables for levels of factors),

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix} \text{ so } \mathbf{x}_i^T = \begin{bmatrix} x_{i1} & \cdots & x_{ip} \end{bmatrix}$$

and

$\boldsymbol{\beta}$ is the $p \times 1$ vector of parameters $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$.

Example 1: Normal linear model

For this model,

$$\mathrm{E}(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}; \qquad Y_i \sim \mathrm{N}(\mu_i, \sigma^2),$$

where $Y_1, \ldots, Y_N$ are independent. Here the link function is the identity function, $g(\mu_i) = \mu_i$. This model is usually written in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where $\mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$ and the e_i's are independent, identically distributed random variables with $e_i \sim \mathrm{N}(0, \sigma^2)$ for $i = 1, \ldots, N$.

Mortality rates

For a large population the probability of a randomly chosen individual dying at a particular time is small. If we assume that deaths from a non-infectious disease are independent events, then the number of deaths Y in a population can be modeled by a Poisson distribution

$$f(y; \mu) = \frac{\mu^y e^{-\mu}}{y!},$$

where $\mu = E(Y)$ is the expected number of deaths in a specified time period.

The parameter μ will depend on the population size, the period of observation and various characteristics of the population (e.g., age, sex and medical history). It can be modeled, for example, by

$$\mu = n\lambda(X^T\boldsymbol{\beta}).$$

where n is the population size and $\lambda(X^T\boldsymbol{\beta})$ is the rate per 100,000 people per year.