Generalized Linear Model

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

$E(y) = b'(\theta)$

$Var(y) = \phi \cdot b''(\theta)$

A linear model $Y = X\boldsymbol{\beta} + \varepsilon$ can be equivalently expressed as

$$\Leftrightarrow \begin{cases} Y|X \sim N(\mu, \sigma^2), \\ \mu = X\boldsymbol{\beta}, \quad \boldsymbol{\beta} \in (-\infty, \infty). \end{cases}$$

The model specifies

- the response variable is continuous and *normally* distributed.
- some function of the mean μ (*the identity function here*) can be written as a linear combination of the covariates.

linear predictor

A generalized linear model (GLM) generalizes normal linear regression models to address a broader class of data structures.

- Instead of being normal, the response variable Y could have any distribution from the **exponential family** distributions.
- The mean $\mu = \mathbb{E}(Y|X)$ may be a more general function of $X\boldsymbol{\beta}$, rather than an identify function.

$$g(\mu) = X\boldsymbol{\beta}$$

Examples

Disease Occurring Rate:

- In the early stages of a disease epidemic, the rate at which new cases occur can often increase exponentially through time.
- We are interested in predicting the number of new cases y_i on day x_i.
- Since y_i is count-valued, we may use the Poisson distribution to model it. $\mu_i = E(y_i)$
- Let μ_i be the expected number of new cases on day x_i. Based on the description, the following model seems reasonable.

$$\mu_i = \beta_0 \exp(\beta_1 x_i)$$

$$\log \mu_i = \underbrace{\log \beta_0}_{\beta_0^*} + \beta_1 x_i = \beta_0^* + \beta_1 x_i$$

Kyphosis Data:

- Children are followed up after corrective spinal surgeries. We are interested in the relationship between clinical covariates and postoperative deforming.
- Binary response: presence or absence of a postoperative deforming (denoted by a binary variable y_i)

$$y_i \sim Bernoulli(\pi_i) \qquad \pi_i = X_i\beta$$

- Assume log odds of deforming is associated with the linear predictor:

$$\log \frac{\pi_i}{1-\pi_i} = X_i\beta$$

The Basics of GLM

We can view the traditional linear model $Y|X \sim N(X\boldsymbol{\beta}, \sigma^2)$ as a combination of three components,

1. a systematic component (or linear predictor):

$$\eta = X\boldsymbol{\beta},$$

2. a random component:

$$Y|X \sim \text{Normal},$$

$$\mathbb{E}(Y|X) = \mu, \ \text{var}(Y|X) = \sigma^2,$$

3. a link function (an identify link):

$$\mu = \eta.$$

- GLM generalizes both the random component and the link function.
- As for the random component, the focus is on distributions in the exponential family, which include many useful special cases such as Normal, Poisson, Gamma, Binomial, etc.
- As for the link function, the focus is to extend the identity link to other *monotone* functions such as reciprocal, log, probit, logit functions, etc. The specific choices depend on real situations.
- *Tip: Always keep in mind: linear regression is a special case of GLM (with normal distribution and identity link)*

Link function

y

$\theta \xrightarrow{b'(\theta)=\mu} \mu \xrightarrow[\eta=g(\mu)]{} \eta = X\beta$

Suppose y has a density from an exponential family:

$$f(y;\theta,\phi) = e^{\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)}.$$

For n observations, $(y_i, x_{i1}, \cdots, x_{ip})$, $i = 1, \cdots, n$

- $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$ is the linear predictor.
- $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is the parameter of interest, and needs to appear somehow in the likelihood function.
- A link function g relates the linear predictor η_i to the mean parameter μ_i:

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

- With a little abuse of notation, sometimes we write $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ to represent entry-wise mapping
- g is required to be *monotone increasing* and *differentiable*.

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) = g^{-1}(\boldsymbol{X}\beta).$$

- It's generally preferred that the image of g is $\mathbb{R}$. (The domain depends on the exponential family.)

$\mu \xrightarrow{g} \eta \in \mathbb{R}$

1. hypothesis
2. range of domain
3. canonical link

Examples of link functions:

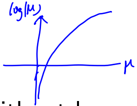
- If μ is unbounded (e.g., Normal distribution), may use identity link

$$\underline{g(\mu) = \mu = \eta}$$

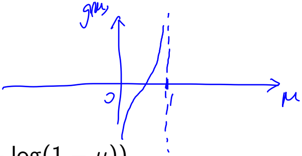
- If μ is positive (e.g., Poisson, Exponential), may use log link

$$\mathbb{R}^+ \; \mu \xrightarrow{g} \eta \; \mathbb{R}$$

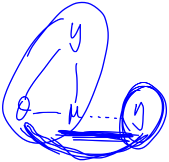
$$g(\mu) = \log(\mu)$$



- If μ is bounded (e.g., Binomial, Bernoulli), without loss of generality, consider $0 < \mu < 1$:
  - logit link: $g(\mu) = \underline{\mathit{logit}(\mu)} = \underline{\log \frac{\mu}{1-\mu}}$;
  - probit link: $g(\mu) = \underline{\Phi^{-1}(\mu);}$
  - complementary log-log link: $g(\mu) = \log(-\log(1-\mu))$.



Canonical Link Functions



Parameter relations:

$$\theta \longleftarrow \mu = \mathbb{E}(y) \longrightarrow \eta$$

Can we connect the natural parameter θ with the linear predictor?

- Canonical Link: the special link function g which makes $\theta = \eta$.
- $g(\mu) = \eta = \theta = b'^{-1}(\mu)$, namely

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

$b'^{-1}(\mu)$

b'^{-1}

- We know b' is strictly increasing and differentiable, so its inverse is a valid link function.

Examples

$f(y) = \frac{1}{\sqrt{2\pi}} e^{\mu y - \frac{\mu^2}{2} - \frac{y^2}{2}}$

$\theta = \mu \quad b(\theta) = \frac{\theta^2}{2}$

$b'(\theta) = \theta \quad b'^{-1}(\mu) = \mu$

$g(\mu) = \mu$

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

$= \frac{1}{y!} e^{-\lambda + y \log \lambda}$

$= \frac{1}{y!} e^{y \cdot \theta - e^\theta}$

$\theta = \log \lambda = \eta$

$\lambda = e^\theta$

$b(\theta) = e^\theta \quad b'(\theta) = e^\theta$

$b'^{-1}(\mu) = \log \mu$

- Normal: Identity link $g(\mu) = \mu$
- Poisson: Log link $g(\lambda) = \log(\lambda)$
- Binary: Logit link $g(\pi) = \log \frac{\pi}{1-\pi}$
- Exponential: Negative reciprocal link $g(\mu) = -1/\mu$

$f(y) = \pi^y (1-\pi)^{1-y}$

$= e^{y \log \pi + (1-y) \log(1-\pi)}$

$= e^{y \log \frac{\pi}{1-\pi} + \log(1-\pi)}$

$= e^{y\theta - \log(1+e^\theta)}$

$\theta = \log \frac{\pi}{1-\pi}$

$\pi = \frac{e^\theta}{1+e^\theta}$

$b(\theta) = \log(1+e^\theta)$

$b'(\theta) = \frac{e^\theta}{1+e^\theta} = \mu$

$b'^{-1}(\mu) = \log \frac{\mu}{1-\mu}$

$f(y) = \lambda \cdot e^{-\lambda x}$

$= e^{-\lambda x + \log \lambda}$

$\theta = -\lambda$

$= e^{x\theta - (-\log(-\theta))}$

$b(\theta) = -\log(-\theta)$

$b'(\theta) = -\frac{1}{\theta}$

$b'^{-1}(\mu) = -\frac{1}{\mu}$

$\theta = -\lambda = \eta$

$-\frac{1}{\mu} = \eta$

GLM Model Fitting

- In GLM of $(\boldsymbol{y}, \boldsymbol{X})$ with a given link function, we can write out the likelihood function as a function of $\boldsymbol{\beta}$
- To estimate $\boldsymbol{\beta}$, we use maximum likelihood (ML) approach
- However, unlike LM, no closed-form MLE for $\boldsymbol{\beta}$
- Need to maximize the log likelihood function numerically
  - Newton-Raphson method
  - Fisher-Scoring method
  - Iteratively reweighted least squares (IRLS) algorithm

Example (Logistic Regression)

Suppose $y_i \sim Bin(1, p_i)$, $i = 1, \ldots, n$, are independent 0/1 indicator responses, and $\boldsymbol{x}_i$ denote a $p \times 1$ vector of predictors for individual i. The log likelihood is as follows

$$\begin{aligned} l(\boldsymbol{y}|\boldsymbol{\beta}) &= \sum_{i=1}^{n} \log\left[p_i^{y_i}(1-p_i)^{(1-y_i)}\right] \\ &= \sum_{i=1}^{n} y_i \log(\frac{p_i}{1-p_i}) - \log(\frac{1}{1-p_i}) \\ &= \sum_{i=1}^{n} (y_i\theta_i - \log(1+e^{\theta_i})). \end{aligned}$$

Choosing the canonical link, the logit link in this case,

$$\eta_i = \theta_i = \log(\frac{p_i}{1-p_i}) = \boldsymbol{x}_i^T\boldsymbol{\beta},$$

which leads to

$$l(\boldsymbol{y}|\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i \boldsymbol{x}_i^T\boldsymbol{\beta} - \log(1+e^{\boldsymbol{x}_i^T\boldsymbol{\beta}})\}.$$

No closed-form MLE!