

# Generalized Linear Model

---

## Why do we need GLM?

Ordinary linear regression predicts the expected value of a given unknown quantity as a linear combination of a set of observed values. This is appropriate when the response variable has a normal distribution. However, these assumptions are inappropriate for some types of response variables.

For example, when the response is not normal, e.g. binary data, the linear model is inappropriate. Sometimes, the output doesn't linearly change with respect to input. Therefore, the ordinary linear regression need to be extended.

## Definition of GLM

Both of OLM and GLM use the **linear predictor**  $\eta = x^\top \beta$  through  $\mu = \mathbb{E}(y)$ .

OLM	GLM
$y \sim \mathcal{N}(\mu, \sigma)$	$f(y; \theta, \phi) = \exp\{\frac{y\theta - b(\theta)}{\phi} + c(y; \phi)\}$
$\mu = \eta$	$\eta = g(\mu)$

where  $\theta$  depends on the **linear predictor**, the **dispersion parameter**  $\phi$  is often *known*, and  $g$  is a monotone **link function**.

The GLM generalized OLM in the aspects of the distribution of response variables and the link between  $\mu$  and  $\eta$  (nonlinear relations between the mean response and the covariates).

## Exponential Family

The GLM extend the distribution of  $y$  from the normal distribution to the exponential family. There are some useful properties:

- Moment-generating function:  $M(t) = \exp\{b(\theta + t\phi) - b(\theta)\}$
- Mean:  $E(Y) = b'(\theta) = \mu$
- Variance:  $\text{var}(Y) = \phi b''(\theta) = \phi b''\{b'^{-1}(\mu)\} = \phi V(\mu)$

where  $V(\mu)$  is the **variance function**.

Note: If  $\phi$  is known, the variance function  $V(\mu)$  and its domain  $\Rightarrow \kappa(\theta) \Rightarrow m(\theta) \Rightarrow$  distribution.

## Link Function $g(\mu)$

The link function provides the relationship between the linear predictor and the mean of the distribution function.

The link function  $g$  should remove restrictions on  $\mu$  by mapping to  $\eta$ :

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

**Canonical link:**  $\eta = \theta = b'^{-1}(\mu)$ . When  $\phi$  is known, the model is a natural exponential family.

## Estimation and Inference

### Discussion of $\beta$

#### Estimation of $\beta$

To estimate  $\beta$ , with Chain rule, we solve the equation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \frac{\partial \eta^T}{\partial \beta} \frac{\partial \theta}{\partial \eta^T} \frac{\partial \ell}{\partial \theta^T} = X^T u(\beta)$$

$\beta$  and  $\theta$  interacts through  $\eta = X^T \beta$ . That is,  $\beta - \eta - \mu - \theta$ .

For the exponential family, the elements of score statistics  $u(\beta)$  and weighted matrix  $W(\beta)$  are

$$u_j = \frac{\partial \theta_j}{\partial \eta_j} \frac{\partial \ell_j(\theta_j)}{\partial \theta_j} = \frac{y_j - \mu_j}{g'(\mu_j) \phi_j V(\mu_j)}$$

$$w_j = \left( \frac{\partial \theta_j}{\partial \eta_j} \right)^2 \frac{\partial^2 \ell_j(\theta_j)}{\partial \theta_j^2} = \frac{1}{g'(\mu_j)^2 \phi_j V(\mu_j)}$$

So, given  $I(\beta) = X(\beta)^T W(\beta) X(\beta)$ , we can find  $\beta$  by iterative weighted least squares method:

$$\hat{\beta} \doteq \beta + I(\beta)^{-1} \frac{\partial \eta^T(\beta)}{\partial \beta} u(\beta)$$

In this case,  $y$  is a possible initial value for  $\mu$  and  $g(y)$  is for  $\eta$ .

#### Inference of $\beta$

Under regularity conditions,

$$\hat{\beta} \sim \mathcal{N}(\beta, (X^T W X)^{-1})$$

### Discussion of $\phi$

At the discussion above, we assume  $\phi_j$  are known. But if they are unknown and have  $\phi_j = a_j \phi$ ,  $\phi$  need to be estimated.

## Estimation of $\phi$

Note that,  $\phi = \text{var}(y_j) / \{a_j V(\mu_j)\}$ .

If  $\beta$  were known,

$$\hat{\phi} = \frac{1}{n} \sum_{j=1}^n \frac{(y_j - \mu_j)^2}{a_j V(\hat{\mu}_j)}$$

If  $\beta$  were unknown,

$$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^n \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

$$P = \frac{1}{\hat{\phi}} \sum_{j=1}^n \frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j) / a_j}$$

## Fit Measure

### Analysis of Deviance

Use scaled deviance:

$$D^* = 2\{l_{max} - l(\hat{\beta})\} \sim \chi_{n-p}^2$$

### Pearson's Statistics

$$P = \frac{1}{\hat{\phi}} \sum_{j=1}^n \frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j) a_j} \sim \chi_{n-p}^2$$

Note: Their distributions depend on the situation.

### Full Model v.s. Reduced Model

For model A and model B:

$$D_A - D_B \sim \chi_{p_A - p_B}^2$$

## References

- 
- [Introduction to Generalized Linear Models](#)
  - [Statistical Models](#)