

GLM - common core

GLM: 3 ingredients

1) $Y_i \sim f(y_i; \theta_i, \phi, w_i)$ of interest

$$L = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i) \right)$$
↙ nuisance

$\mu_i = E(Y_i) = b'(\theta_i)$
 $\sigma_i^2 = \text{Var}(Y_i) = \underbrace{b''(\theta_i)}_{\phi} \cdot w_i$

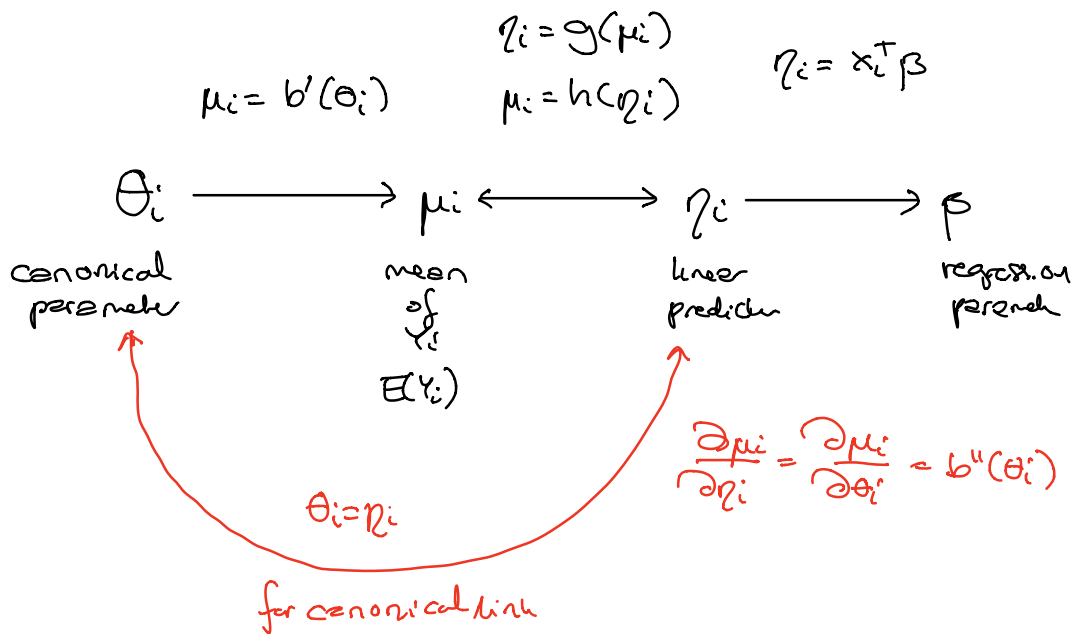
$V(\mu_i) \leftarrow$ variance function

2) $\eta_i = X_i^T \beta$, $\Sigma = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \end{bmatrix} \leftarrow \text{full rank } (p)$

3) $\eta_i = g(\mu_i) \Leftrightarrow \mu_i = h(\eta_i)$ - require one-to-one
link response - twice differentiable

- N: $\eta_i = \mu_i$
- bin: $\mu_i = p_i (n_i = 1)$: $\eta_i = \ln\left(\frac{p_i}{1-p_i}\right)$
- Poisson: $\mu_i = \lambda_i$: $\eta_i = \ln \lambda_i$
- gamma: $\eta_i = \ln \mu_i$, $-\frac{1}{\mu_i}$ (canonical link)

Likelihood inference



Loglikelihood

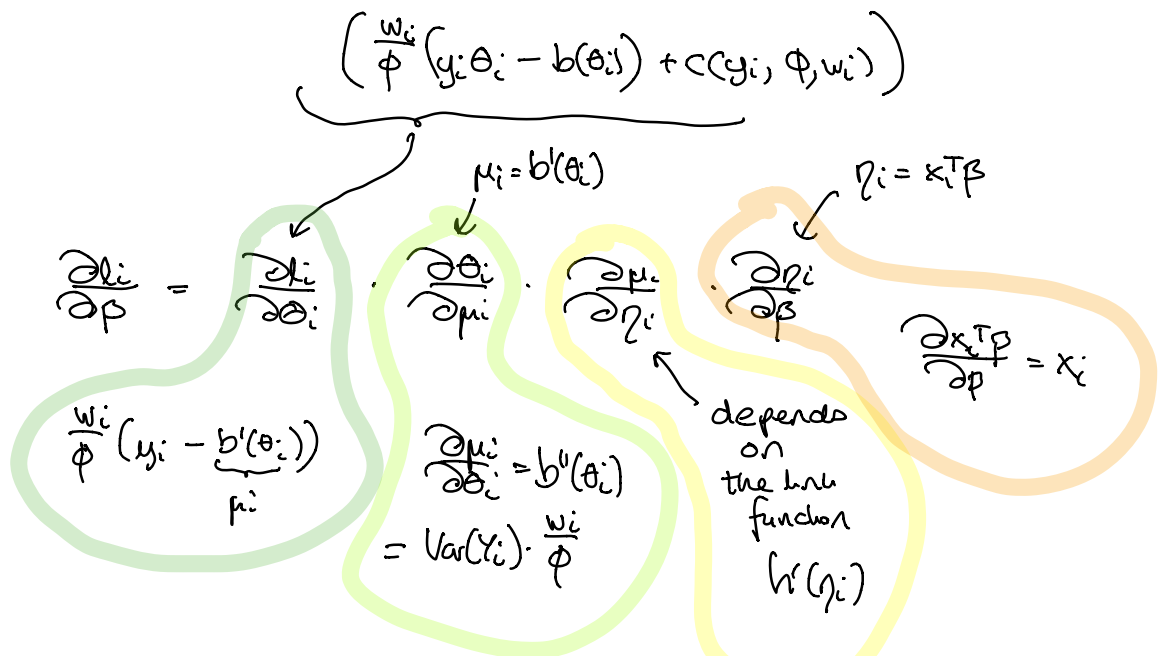
canonical link

$$\sum_{i=1}^n y_i \eta_i = \sum_{i=1}^n y_i x_i^T \beta = \sum_j \beta_j \sum_i x_{ij} y_i$$

$$l(\beta) = \sum_{i=1}^n l_i(\beta) = \sum_{i=1}^n \left(\frac{w_i}{\phi} (y_i \theta_i - b(\theta_i)) + c(y_i, \phi, w_i) \right)$$

Score function $S(\beta) = \frac{\partial l}{\partial \beta}$

$$S(\beta) = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta} = \sum_{i=1}^n s_i(\beta)$$



$$\frac{\partial l_i}{\partial \beta} = (y_i - \mu_i) \cdot \frac{w_i}{\phi} \cdot \frac{\phi}{w_i} \cdot \frac{1}{\text{Var}(Y_i)} \cdot h'(\eta_i) \cdot x_i$$

$$s(\beta) = \sum_{i=1}^n \frac{(y_i - \mu_i) \cdot x_i \cdot h'(\eta_i)}{\text{Var}(Y_i)}$$

canonical link $h'(\eta_i) = b''(\theta_i) = \text{Var}(Y_i) \frac{w_i}{\phi}$

$$s(\beta) = \sum (y_i - \mu_i) x_i \frac{w_i}{\phi}$$

when $w_i = \phi = 1$ as for Poisson and Bernoulli, then likelihood eq. or

$$\sum x_{ij} y_i = \sum x_{ij} \mu_i$$

the likelihood eq. equate the sufficient statistics for the model param. to their expected value

added!

Observe: $s(\beta)$ depends on the distr. of Y_i only through μ and $\text{Var}(Y_i)$.

Expected Fisher information matrix $F(\beta)$

$$F(\beta) = \text{Cov}(s(\beta)) = \sum_{i=1}^n \underbrace{\text{Cov}(s_i(\beta))}_{F_i(\beta)} \quad \text{and} \quad \text{Cov}(s_i(\beta)) = E(s_i(\beta)s_i(\beta)^T) \\ \text{since } E(s_i(\beta)) = 0$$

\uparrow
 indep. $i=1, \dots, n$
 obs

because: $E(s_i(\beta)) = E\left(\frac{(Y_i - \mu_i) x_i \cdot h'(\eta_i)}{\text{Var}(Y_i)}\right) = \frac{x_i h'(\eta_i)}{\text{Var}(Y_i)} \left[\underbrace{E(Y_i)}_{\mu_i} - \mu_i \right] = 0$

We will look at element (h, l) of $F_i(\beta)$, and use

$$\frac{\partial l_i}{\partial \beta_h} = \frac{(Y_i - \mu_i) \cdot x_{ih} \cdot h'(\eta_i)}{\text{Var}(Y_i)}$$

$$F_i(\beta)_{[h, l]} = E\left(\frac{\partial l_i}{\partial \beta_h} \cdot \frac{\partial l_i}{\partial \beta_l}\right)$$

$$= E\left[\frac{(Y_i - \mu_i) \cdot x_{ih} \cdot h'(\eta_i)}{\text{Var}(Y_i)} \cdot \frac{(Y_i - \mu_i) \cdot x_{il} \cdot h'(\eta_i)}{\text{Var}(Y_i)}\right]$$

$$= \frac{h'(\eta_i)^2 \cdot x_{ih} \cdot x_{il}}{\text{Var}(Y_i)^2} \underbrace{E\left[(Y_i - \mu_i)^2\right]}_{\text{Var}(Y_i)} = \frac{x_{ih} \cdot x_{il} \cdot [h'(\eta_i)]^2}{\text{Var}(Y_i)}$$

$$F(\beta)_{[h, l]} = \sum_{i=1}^n \frac{x_{ih} \cdot x_{il} \cdot [h'(\eta_i)]^2}{\text{Var}(Y_i)}$$

$$\underline{F(\beta) = X^T W X} \quad \text{where } W = \text{diag}\left(\frac{h'(\eta_i)^2}{\text{Var}(Y_i)}\right)$$

Alt: $\left. \begin{array}{l} D = \text{diag}(h'(\eta_i)) \\ \Sigma = \text{diag}(\text{Var}(Y_i)) \end{array} \right\} D \Sigma^{-1} D = W$

When canonical link: $H(\beta) = -\frac{\partial \eta}{\partial \beta^T}$
↓
 $E(H(\beta)) = H(\beta)$
F(β)

PARAMETER ESTIMATION:

→ can be done separately for β and ϕ
(see notes on orthogonal parameters - advanced topic)

β is found by solving $S(\beta) = 0$ using the
Fisher scoring ← need formulas for $S(\beta)$ and $F(\beta)$
↓
we get Iterative Reweighted Least Squares
IRWLS
method