GLMs

Thomas Achia

3/24/2022

Generalized Linear models

The Exponential family

The pdf of a random varible Y from the exponential family of distributions:

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]$$

Example. **The Exponential distribution**

$$f(y;\theta) = \theta\exp(-\theta y) = \exp\left[-\theta y + \ln\theta\right]$$
$$a(y) = y; b(\theta) = -\theta; c(\theta) = \ln\theta; d(y) = 0$$

Example. **The Bernoulli distribution**

$$f(y;\theta) = \theta^y(1-\theta)^{1-y} = \exp\left[\ln\left\{\theta^y(1-\theta)^{1-y}\right\}\right] = \exp\left[y\ln\theta + (1-y)\ln(1-\theta)\right]$$

Thus

$$f(y;\theta) = \exp\left[y\ln\left(\frac{\theta}{1-\theta}\right) + \ln(1-\theta)\right]$$

Here

$$a(y) = y; b(\theta) = \ln\left(\frac{\theta}{1-\theta}\right) = \text{logit}(\theta); c(\theta) = \ln(1-\theta); d(y) = 0$$
$$\text{logit} \equiv \text{Canonical link for the Binomial distribution}$$

Note

If $a(y) = y$ then the pdf of Y is in Canonical form and $b(\theta)$ is the canonical link function of $f(y;\theta)$

Example. **The Poisson distribution**

$$f(y;\theta) = \frac{e^{-\theta}\theta^y}{y!} = \exp\left[-\theta + y\ln\theta - \ln y!\right] = \exp\left[y\ln\theta - \theta - \ln y!\right]$$

Here

$$a(y) = y; b(\theta) = \ln\theta; c(\theta) = -\theta; d(y) = -\ln y!$$

$$\text{Natural logarithm} \equiv \text{Canonical link for the Poisson distribution}$$

Example. **The** $\Gamma(\alpha,\beta)$ distribution

$$Y \sim \text{Gamma}(\alpha, \beta)$$
$$f(y;\alpha;\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-y/\beta} y^{\alpha-1}$$

$$\begin{aligned}
f(y;\alpha;\beta) &= \exp\left[-\frac{y}{\beta} + (\alpha-1)\ln y - \ln\Gamma(\alpha) - \alpha\ln\beta\right] \\
&= \exp\left[\frac{-y}{\beta} - \alpha\ln\beta + (\alpha-1)\ln y - \ln\Gamma(\alpha)\right] \\
&= \exp\left[\left(\frac{y(-\alpha\beta)}{\alpha\beta^2} - \frac{\alpha^2\beta^2\ln\beta}{\alpha\beta^2}\right) + (\alpha-1)\ln y - \ln\Gamma(\alpha)\right] \\
&= \exp\left[\left(\frac{y(-\alpha\beta) - \alpha^2\beta^2\ln\beta}{\alpha\beta^2}\right) + (\alpha-1)\ln y - \ln\Gamma(\alpha)\right]
\end{aligned}$$

Then

$$(-\alpha\beta) = \theta \Rightarrow \beta = \frac{-\theta}{\alpha} \Rightarrow \ln\beta = \ln(-\theta) - \ln\alpha$$
$$f(y;\alpha;\beta) = \exp\left[\left(\frac{y\theta - \theta^2[\ln(-\theta) - \ln\alpha]}{\alpha\beta^2}\right) + (\alpha-1)\ln y - \ln\Gamma(\alpha)\right]$$

Task

Show that the following belong to the exponential family

1. The pareto distribution:
$$f(y;\theta) = \theta y^{-\theta-1}$$
2. The Negative Binomial distribution:
$$f(y;\theta) = \begin{pmatrix} y+r-1 \\ r-1 \end{pmatrix} \theta^r (1-\theta)^y.$$

Example. **The Normal distribution**

$$\begin{aligned}
f(y;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(y-\mu)^2 \\
&= \exp\left[-\left(\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right) - \frac{1}{2}\ln 2\pi\sigma^2\right] \\
&= \exp\left[\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln 2\pi\sigma^2\right]
\end{aligned}$$

$$a(y) = y; b(\theta) = \frac{\mu}{\sigma^2}; c(\theta) = -\frac{\mu^2}{2\sigma^2}$$

The Exponential dispersion family (IN CANONICAL FORM)

$$f(y;\theta) = \exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y;\phi)\right]$$

The mean and variance for the Exponential dispersion family

We want to prove that if the pdf of Y belongs to the Exponential dispersion family then the mean and vaiance of Y are respectively

$$E(Y) = b'(\theta) \text{ and } \mathrm{Var}(Y) = \phi V(\mu),$$

where $V(\mu) = b''(\theta)$.

The mean of a random variable belonging to the exponential family

If the pdf of Y belongs to the the Exponential dispersion family then

$$\int_y f(y; \theta, \phi) dy = 1$$

since $f(y; \theta, \phi)$ is a pdf.

Differentiating this expression wrt θ we get

$$\frac{\partial}{\partial\theta} \int_Y f(y; \theta, \phi) dy = \frac{\partial}{\partial\theta}(1) = 0 \Rightarrow \int_Y \frac{\partial}{\partial\theta} \left[f(y; \theta, \phi)\right] dy = 0$$

Then

$$\begin{aligned} 0 &= \int_Y \frac{\partial}{\partial\theta} \left[f(y; \theta, \phi)\right] dy \\ &= \int_Y \frac{\partial}{\partial\theta} \left[\exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y; \phi)\right]\right] dy \\ &= \int_Y \left[\frac{y - b'(\theta)}{\phi}\right] \left[\exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y; \phi)\right]\right] dy \end{aligned}$$

Thus

$$\begin{aligned} 0 &= \int_Y \left[\frac{y - b'(\theta)}{\phi}\right] f(y; \theta; \phi) dy \\ &= \int_Y \left[\frac{y}{\phi}\right] f(y; \theta; \phi) dy - \int_Y \left[\frac{b'(\theta)}{\phi}\right] f(y; \theta; \phi) dy \\ &= \left[\frac{1}{\phi}\right] \underbrace{\int_Y y f(y; \theta; \phi) dy}_{E(Y)} - \left[\frac{b'(\theta)}{\phi}\right] \int_Y f(y; \theta; \phi) dy \end{aligned}$$

$$\left[\frac{1}{\phi}\right] E(Y) - \left[\frac{b'(\theta)}{\phi}\right] = 0 \Rightarrow \boxed{\mu = E(Y) = b'(\theta)}$$

The Variance of a distribution in the exponential dispersion family

For the variance, we take a second derivative of the expression

$$\int_y f(y; \theta, \phi) dy = 1$$

Taking the derivative of

$$\int_Y \left[\frac{y - b'(\theta)}{\phi}\right] \left[\exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y; \phi)\right]\right] dy = 0$$

Then

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta} \int_Y \left[\frac{y - b'(\theta)}{\phi}\right] f(y; \theta, \phi)) dy \\ &= \int_Y \left[\frac{0 - b''(\theta)}{\phi}\right] f(y; \theta, \phi) dy + \int_Y \left[\frac{y - b'(\theta)}{\phi}\right]^2 f(y; \theta, \phi) dy \end{aligned}$$

which simplifies to

$$-\left[\frac{b''(\theta)}{\phi}\right]+E\left[\frac{y-b'(\theta)}{\phi}\right]^2=0$$

Using the relationship

$$E(A^2)-[E(A)]^2=Var(A)\Rightarrow E(A^2)=Var(A)+[E(A)]^2$$

we get

$$-\left[\frac{b''(\theta)}{\phi}\right]+E\left[\frac{y-b'(\theta)}{\phi}\right]^2=-\left[\frac{b''(\theta)}{\phi}\right]+\underbrace{Var\left[\frac{y-b'(\theta)}{\phi}\right]}_{\left(\frac{1}{\phi}\right)^2 Var(Y)}+\underbrace{\left[E\left[\frac{y-b'(\theta)}{\phi}\right]\right]^2}_{0}=0.$$

Thus

$$-\left[\frac{b''(\theta)}{\phi}\right]+\left(\frac{1}{\phi}\right)^2 Var(Y)=0\Rightarrow \boxed{Var(Y)=\phi b''(\theta)\equiv\phi V(\mu)}$$

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

Example. **The Poisson distribution**

$$f(y;\lambda)=\frac{e^{-\lambda}\lambda^y}{y!}=\exp(y\ln\lambda-\lambda-\ln y!)=\exp\left[\frac{y\ln\lambda-\lambda}{1}-\ln y!\right]$$

$$\theta=\ln\lambda\Rightarrow\lambda=e^\theta;b(\theta)=\lambda;\phi=1;c(y;\phi)=\ln y!$$

$$\theta=\ln\lambda\Rightarrow\lambda=e^\theta\Rightarrow\frac{\partial\lambda}{\partial\theta}=e^\theta$$

Taking derivative of $b(\theta)$ with respect to θ we get

$$b'(\theta)=\frac{\partial}{\partial\theta}[b(\theta)]=\frac{\partial}{\partial\theta}[\lambda]=\frac{\partial\lambda}{\partial\theta}=e^\theta=\lambda.$$

Thus

$$b'(\theta)=\lambda=E(Y)$$

Taking a second derivative of $b(\theta)$ with respect to θ we get

$$b''(\theta)=\frac{\partial}{\partial\theta}[b'(\theta)]=\frac{\partial}{\partial\theta}[\lambda]=\frac{\partial\lambda}{\partial\theta}=e^\theta=\lambda$$

Thus

$$b''(\theta)=\lambda\Rightarrow\boxed{\mathrm{Var}(Y)=\phi b''(\theta)=1\times\lambda=\lambda}$$

Maximum Likelihood Estimation for generalized linear models

Let $Y_1,\ldots,Y_n$ be a random sample from the exponential dispersion family. That is,

$$f(y_i;\theta)=\exp\left[\frac{y_i\theta-b(\theta)}{\phi}+c(y;\phi)\right]$$

For generalized linear models, we assume the following:

1. The random variable Y have a pdf $f(y_i;\theta)$ belonging to the exponential dispersion family;
2. The relation between the Y and set of independent variable $X_1,\ldots,X_k$ is given by

$$\eta_i=g(\mu_i)=\beta_0+\beta_1X_{1i}+\ldots+\beta_kX_{ki}=\boldsymbol{X}_i'\boldsymbol{\beta},$$

where $g(.)$ is referred to as a link function that links the systematic to the deterministic part of the more and where $\mu_i=E(Y_i)$.

The Likelihood

The Likelihood function is then:

$$L(\theta,\phi;\boldsymbol{y}) = \prod_{i=1}^{n} f(y_i;\theta) = \prod_{i=1}^{n} \exp\left[\frac{y_i\theta - b(\theta)}{\phi} + c(y;\phi)\right] = \exp\sum_{i=1}^{n}\left[\frac{y_i\theta - b(\theta)}{\phi} + c(y;\phi)\right]$$

and the log-likelihood function is:

Then the log-likelihood function is:

$$\ell(\theta,\phi;\boldsymbol{y}) = \sum_{i=1}^{n}\left[\frac{y_i\theta - b(\theta)}{\phi} + c(y;\phi)\right] = \sum_{i=1}^{n} \ell_i,$$

where

$$\ell_i = \left[\frac{y_i\theta - b(\theta)}{\phi} + c(y;\phi)\right]$$

To find the MLE we seek a statistic that satisfies:

$$S(\theta;\boldsymbol{y}) = \frac{\partial \ell(\theta,\phi;\boldsymbol{y})}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\sum_{i=1}^{n}\ell_i = \sum_{i=1}^{n}\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = 0 \Rightarrow \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = 0.$$

$$\boxed{\eta_i = g(\mu_i)} = \boldsymbol{X}_i\boldsymbol{\beta} \Rightarrow \mu_i = g^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}).\boxed{\frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}g(\mu_i) = g'(\mu_i)}.\boxed{\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{X}_i\boldsymbol{\beta} = \boldsymbol{X}_i.}$$

Now

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \frac{\partial \ell_i}{\partial \theta}\times\frac{\partial \theta}{\partial \boldsymbol{\beta}} = \frac{\partial \ell_i}{\partial \theta}\times\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\times\frac{\partial \theta}{\partial \boldsymbol{\mu}_i} = \frac{\partial \ell_i}{\partial \theta}\times\frac{\partial \mu_i}{\partial \eta_i}\times\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}\times\frac{\partial \theta}{\partial \mu_i} = 0.$$

This becomes

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \underbrace{\left[\frac{y_i - b'(\theta)}{\phi}\right]}_{\frac{\partial \ell_i}{\partial \theta}} \times \left[\frac{1}{\frac{\partial \eta_i}{\partial \mu_i}}\right] \times \underbrace{\boldsymbol{X}_i'}_{\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{X}_i\boldsymbol{\beta}} \times \left[\frac{1}{\frac{\partial \mu_i}{\partial \theta}}\right] = \left[\frac{y_i - b'(\theta)}{\phi}\right]\times\left[\frac{1}{g'(\mu_i)}\right]\times\boldsymbol{X}_i'\times\left[\frac{1}{\frac{\partial b'(\theta)}{\partial \theta}}\right] = \left[\frac{\boldsymbol{X}_i'}{g'(\mu_i)}\right]\left[\frac{y_i - b'(\theta)}{\phi b''(\theta)}\right] = \left[\frac{\boldsymbol{X}_i'(y}{g'(\mu_i}\right.$$

The estimating equation

That is,

$$S(\theta;\boldsymbol{y}) = \sum_{i=1}^{n}\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{\boldsymbol{X}_i'}{g'(\mu_i)}\left[\frac{y_i - \mu_i}{\mathrm{Var}(y_i)}\right] = 0.$$

$$\boxed{S(\theta;\boldsymbol{y}) = \sum_{i=1}^{n}\frac{\boldsymbol{X}_i'}{g'(\mu_i)}\left[\frac{y_i - \mu_i}{\mathrm{Var}(y_i)}\right] = 0.} \equiv \text{Estimating equation}$$

$$\text{Systematic component } \equiv \eta_i = g(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta} \Rightarrow \mu_i = g^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})$$

Example. **For the Normal distribution**

$$\text{The pdf: } f(y;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp -\frac{1}{2\sigma^2}(y-\theta)^2 = \exp\left\{\frac{y\theta - \theta^2/2}{\sigma^2} - \frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln 2\pi\sigma^2\right]\right\}$$

$$b(\theta) = \frac{\theta^2}{2} \Rightarrow b'(\theta) = \mu = \theta \Rightarrow b''(\theta) = 1;\ \phi = \sigma^2;\ c(y;\phi) = -\frac{1}{2}\left[\frac{y^2}{\sigma^2} + \ln 2\pi\sigma^2\right]$$

Here

$$\text{Var}(y) = \phi V(\mu_i) = \phi b''(\theta) = \sigma^2 \times 1 = \sigma^2, \quad \text{where } V(\mu_i) = 1 \text{ in this case.}$$

and

$$g(\mu) = \mu \text{ since the distribution is in Canonical form and has the identity link function.} = \boldsymbol{X\beta} \Rightarrow \boxed{\mu = \boldsymbol{X\beta}}$$

This implies that

$$g'(\mu) = \frac{\partial}{\partial \mu} g(\mu) = 1.$$

Hence the estimating equation is:

$$S(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{g'(\mu_i)} \left[\frac{y_i - \mu_i}{\text{Var}(y_i)} \right] = \sum_{i=1}^{n} \boldsymbol{X}_i' \left[\frac{y_i - \mu_i}{\sigma^2} \right] = 0 \Rightarrow \boxed{\sum_{i=1}^{n} \boldsymbol{X}_i' (y_i - \boldsymbol{X\beta}) = 0} \Rightarrow \boldsymbol{X'y} - \boldsymbol{X'X\beta} = 0$$

That is,

$$\boldsymbol{X'X\beta} = \boldsymbol{X'y} \Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1} \boldsymbol{X'y} \equiv \text{An analytical solution}$$

Example. **The Poisson distribution**

$$\text{The pdf: } f(y; \theta) = \frac{e^{-\lambda} \lambda^y}{y!} = \exp\{y \ln \lambda - \lambda - \ln y!\} = \exp\left\{ \frac{y \ln \lambda - \lambda}{1} - \ln y! \right\}$$

$$\theta = \ln \lambda \Rightarrow \frac{\partial \theta}{\partial \lambda} = \frac{1}{\lambda} = e^{-\theta}; b(\theta) = \lambda \Rightarrow b'(\theta) = \frac{\partial \lambda}{\partial \theta} = e^{\theta} = \mu \Rightarrow b''(\theta) = e^{\theta}; \ \phi = 1; \ c(y; \phi) = -\ln y!$$

Here

$$\text{Var}(y) = \phi V(\mu_i) = \phi b''(\theta) = 1 \times e^{\theta} = \lambda, \quad \text{where } V(\mu_i) = \lambda \text{ in this case.}$$

and

$$g(\mu) = \ln \mu \text{ since the canonical form is the natural logarithm function.} = \boldsymbol{X\beta} \Rightarrow \boxed{\ln \mu = \boldsymbol{X\beta}} \Rightarrow \ln \lambda = \boldsymbol{X\beta} \Rightarrow \lambda = \exp(\boldsymbol{X\beta})$$

This implies that

$$g'(\mu) = \frac{\partial}{\partial \mu} g(\mu) = \frac{\partial}{\partial \mu} \ln \mu = \frac{1}{\mu} = \frac{1}{\lambda} = \exp(-\boldsymbol{X\beta}).$$

Hence the estimating equation is:

$$S(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{g'(\mu_i)} \left[\frac{y_i - \mu_i}{\text{Var}(y_i)} \right] = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{\frac{1}{\lambda}} \left[\frac{y_i - \lambda}{\lambda} \right] = 0 \Rightarrow \sum_{i=1}^{n} \boldsymbol{X}_i' (y_i - \lambda) = 0 \Rightarrow \boxed{\sum_{i=1}^{n} \boldsymbol{X}_i' \left(y_i - e^{X\boldsymbol{\beta}} \right) = 0}$$

$$\sum_{i=1}^{n} \boldsymbol{X}_i' y_i = \sum_{i=1}^{n} \boldsymbol{X}_i' e^{X\boldsymbol{\beta}} \Rightarrow \boldsymbol{X'y} = \sum_{i=1}^{n} \boldsymbol{X}_i' e^{X\boldsymbol{\beta}} \equiv \text{There is no analytical solution; We resort to numerical approximation}$$

No analytical solution exists

Example. **Bernoulli distribution**

$$f(y; \pi) = \pi^y (1 - \pi)^{1-y} = \exp\left[\frac{y \boxed{\ln\left(\frac{\pi}{1 - \pi} \right)} + \ln(1 - \pi)}{1} \right]$$

$$\theta = \ln\left(\frac{\pi}{1-\pi}\right) \Rightarrow \pi = \frac{e^\theta}{1+e^\theta} \Rightarrow \frac{\partial \pi}{\partial \theta} = \frac{(1+e^\theta)e^\theta - (e^\theta)^2}{(1+e^\theta)^2} = \frac{e^\theta}{(1+e^\theta)^2} = \frac{\frac{\pi}{1-\pi}}{(1+\frac{\pi}{1-\pi})^2} = \pi(1-\pi).$$

$$b(\theta) = -\ln(1-\pi) \Rightarrow b'(\theta) = \frac{\partial}{\partial \theta}[-\ln(1-\pi)] = \frac{\partial}{\partial \pi}[-\ln(1-\pi)]\frac{\partial \pi}{\partial \theta} = \left(\frac{1}{1-\pi}\right)\frac{\partial \pi}{\partial \theta} = \left(\frac{1}{1-\pi}\right)\pi(1-\pi) = \pi$$

$$\boxed{\mu_i = b'(\theta) = \pi}$$

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

$$\phi = 1$$

$$\text{Var}(y) = \phi V(\mu) = \phi b''(\theta) = \pi(1-\pi)$$

$$S(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{g'(\mu_i)}\left[\frac{y_i - \mu_i}{\text{Var}(y_i)}\right] = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{g'(\mu_i)}\left[\frac{y_i - \pi}{\pi(1-\pi)}\right] = 0.$$

$$\boxed{g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X\beta}} \Rightarrow g'(\mu_i) = \frac{\partial}{\partial \pi}\ln\left(\frac{\pi}{1-\pi}\right) = \left(\frac{\pi}{1-\pi}\right)^{-1} \times \frac{(1-\pi)+\pi}{(1-\pi)^2} = \frac{1}{\pi(1-\pi)}$$

Hence

$$S(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{X}_i'\left[y_i - \pi\right] = 0.$$

$$\boxed{g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X\beta}} \Rightarrow \pi = \frac{e^{\boldsymbol{X\beta}}}{1+e^{\boldsymbol{X\beta}}}$$

$$S(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{X}_i'\left[y_i - \frac{e^{\boldsymbol{X\beta}}}{1+e^{\boldsymbol{X\beta}}}\right] = 0.$$

$$\boldsymbol{X}'\boldsymbol{y} = \sum_{i=1}^{n} \boldsymbol{X}_i'\left[\frac{e^{\boldsymbol{X\beta}}}{1+e^{\boldsymbol{X\beta}}}\right] \equiv \text{The estimating equation with no analytical solution}$$

Note

A numerical solution to the problem is required.

Possible soultions: We will describe the following three methods:

1. Newton-Raphson Method
2. Fisher-scoring Method
3. Iteratively Re-weighted Least Squares

Fisher-scoring Method

The score function:

$$\boxed{S(\boldsymbol{\beta}; \boldsymbol{y}) = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\boldsymbol{X}_i'}{g'(\mu_i)}\left[\frac{y_i - \mu_i}{\text{Var}(y_i)}\right] = 0.} \equiv \text{Estimating equation}$$

The Information:

$$I(\boldsymbol{\beta}; \boldsymbol{y}) = -\frac{\partial}{\partial \boldsymbol{\beta}} S(\boldsymbol{\beta}; \boldsymbol{y}) = -\underbrace{H(\boldsymbol{\beta}; \boldsymbol{y})}_{Hessian}$$

We want to find the MLE of $\boldsymbol{\beta}$ using the Taylor series expansion of $S(\boldsymbol{\beta};\boldsymbol{y})$:

$$f(x) = f(a) + (x-a)f^{(1)}(a) + \frac{(x-a)^2}{2!}f^{(2)}(a) + \ldots \approx f(x) = f(a) + (x-a)f^{(1)}(a)$$

$$\underbrace{S(\boldsymbol{\beta};\boldsymbol{y})}_{p\times 1} \approx \underbrace{S(\boldsymbol{\beta}^*;\boldsymbol{y})}_{p\times 1} + \underbrace{\left[\frac{\partial}{\partial\boldsymbol{\beta}}S(\boldsymbol{\beta};\boldsymbol{y})\right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}}_{p\times p}\underbrace{(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}_{p\times 1}$$

$$\underbrace{S(\boldsymbol{\beta};\boldsymbol{y})}_{p\times 1} \approx \underbrace{S(\boldsymbol{\beta}^*;\boldsymbol{y})}_{p\times 1} - \underbrace{I(\boldsymbol{\beta}^*;\boldsymbol{y})}_{p\times p}\underbrace{(\boldsymbol{\beta}-\boldsymbol{\beta}^*)}_{p\times 1}$$

$$S(\boldsymbol{\beta};\boldsymbol{y}) \approx \underbrace{S(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})}_{0} - I(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{MLE}) \Rightarrow S(\boldsymbol{\beta};\boldsymbol{y}) \approx -I(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{MLE}) \Rightarrow (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{MLE}) \approx -I^{-1}(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})S(\boldsymbol{\beta};\boldsymbol{y})$$

$$(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{MLE}) \approx -I^{-1}(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})S(\boldsymbol{\beta};\boldsymbol{y}) \Rightarrow \hat{\boldsymbol{\beta}}_{MLE} = \boldsymbol{\beta} + I^{-1}(\hat{\boldsymbol{\beta}}_{MLE};\boldsymbol{y})S(\boldsymbol{\beta};\boldsymbol{y})$$

$$\hat{\boldsymbol{\beta}}^{r+1}_{MLE} = \hat{\boldsymbol{\beta}}^{r}_{MLE} + I^{-1}(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y})S(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y})$$

$$\boxed{\hat{\boldsymbol{\beta}}^{r+1}_{MLE} = \hat{\boldsymbol{\beta}}^{r}_{MLE} - H^{-1}(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y})S(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y})}$$

$$H^{-1}(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y}) = -I^{-1}(\hat{\boldsymbol{\beta}}^{r}_{MLE};\boldsymbol{y}) \equiv \text{Hessian} = \text{Negative the information}$$

Examples of Generalized linear models

$$\text{logit}(p_i) = \boldsymbol{X}_i\boldsymbol{\beta} \Rightarrow \ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X}_i\boldsymbol{\beta} \Rightarrow \hat{p}_i = \frac{e^{\boldsymbol{X}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}_i\boldsymbol{\beta}}}$$

Example

The result

$$\text{logit}(p_i) = -60.72 + 34.27\ conc \Rightarrow p_i = \frac{e^{-60.72+34.27\ conc}}{1+e^{-60.72+34.27\ conc}}$$

$$[\hat{p}_i]_{conc=2} = \frac{e^{-60.72+34.27\times 2}}{1+e^{-60.72+34.27\times 2}}$$

$$[\hat{p}_i]_{conc=1.5} = \frac{e^{-60.72+34.27\times 1.5}}{1+e^{-60.72+34.27\times 1.5}}$$

$$\text{Levels of B} \equiv b_1, b_2, \ldots, b_b$$

$$logit(p_i) = \beta_0 + \beta_2 B_2 + \ldots + \beta_p B_p$$

$$\text{Levels of B} \equiv b_1,\ b_2$$

$$logit(p_i) = \beta_0 + \beta_2 B_{2i},$$

$$B_2 = \begin{cases} 1 & \text{if the } i-\text{th individual belongs to group2} \\ 0 & \text{if the } i-\text{th individual belongs to group1} \end{cases}$$

For group 1 ($B_{2i} = 0$)

$$logit(p_{1i}) = \ln\left(\frac{p_{1i}}{1-p_{1i}}\right) = \ln O_1 = \beta_0 + \beta_2 \times 0 = \beta_0 \Rightarrow \ln O_1 = \beta_0 \Rightarrow O_1 = \exp(\beta_0)$$

For group 2 ($B_{2i} = 1$)

$$logit(p_{2i}) = \ln\left(\frac{p_{2i}}{1-p_{2i}}\right) = \ln O_2 = \beta_0 + \beta_2 \times 1 = \beta_0 + \beta_2 \Rightarrow \ln O_2 = \beta_0 + \beta_2 \Rightarrow O_2 = \exp(\beta_0 + \beta_2)$$

$$OR = \frac{O_2}{O_1} = \frac{\exp(\beta_0 + \beta_2)}{\exp(\beta_0)} = \exp \beta_2$$