Generalized Linear Models:

A subclass of nonlinear models with a "linear" component

<u>Data:</u>

$$
\begin{array}{c}
(Y_1, \mathbf{X}_1^T) \\
(Y_2, \mathbf{X}_2^T) \\
\vdots \\
(Y_n, \mathbf{X}_n^T) \\
\nearrow \qquad \nwarrow
\end{array}
$$

response variable — explantory or regression, variables

where

$$\mathbf{X}_j^T = (X_{1j}, X_{2j}, \ldots, X_{pj})$$

describe the conditions under which the response Y_j was obtained.

Basic Features:

1. $Y_1, Y_2, \cdots, Y_n$ are independent

2. Y_j has a distribution in the exponential family of distributions, $j = 1, 2, \cdots, n$.

   The joint likelihood has the form (canonical form):

$$L(\boldsymbol{\theta}, \phi; \mathbf{Y}) = \pi_{j=1}^{n} f(Y_j; \theta_j, \phi)$$

   where

$$f(Y_j; \theta_j \phi) = \exp\{\frac{Y_j\theta_j - b(\theta_j)}{a(\phi)}\}$$

   for some functions

$$a(), b(), \quad \text{and} \quad c().$$

   Here, θ_j is called a canonical parameter.

3. (Systematic part of the model)

There is a link function $h()$ that links the conditional mean of Y_j given $\mathbf{X}_j = (X_{ij}, \ldots, X_{pj})^T$ to a linear function of unknown parameters, e.g.

$$h(E(Y_j|\mathbf{X}_j)) = \mathbf{X}_j^T \boldsymbol{\beta} = \beta_1 X_{1j} + \ldots + \beta_p X_{pj}$$

Members of the exponential family of distributions available in

S-PLUS: glm() function
SAS: PROC GENMOD

Normal distribution: $Y \sim N(\mu, \sigma^2)$

$$f(Y;\theta,\phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(Y-\mu)^2}{2\sigma^2}\right\}$$

$$= \exp\left\{\frac{Y\mu - \frac{\mu^2}{2}}{\sigma^2} \underbrace{- \frac{1}{2}\left[\frac{Y^2}{\sigma^2} + \log(2\pi\sigma^2)\right]}\right\}$$

$$\uparrow C(Y,\phi)$$

Here

$$E(Y) = \mu \equiv \theta \qquad\qquad b(\theta) = \frac{\theta^2}{2}$$

and

$$Var(Y) = \sigma^2 = \phi \qquad\qquad a(\phi) = \phi$$

Binomial Distribution: $Y \sim Bin(n, \pi)$

$$f(Y;\theta,\phi) = \binom{n}{Y}\pi^Y(1-\pi)^{n-Y}$$

$$= \binom{n}{Y}\left(\frac{\pi}{1-\pi}\right)^Y (1-\pi)^n$$

$$= \exp\left\{Y\log\left(\frac{\pi}{1-\pi}\right) + n\log(1-\pi) + \underline{\log\binom{n}{Y}}\right\}$$

$$\uparrow C(Y,\phi).$$

Here

$$\theta = \log\left(\frac{\pi}{1-\pi}\right) \Rightarrow \pi = \frac{e^\theta}{1+e^\theta}$$

and

$$b(\theta) = n\log(1-\pi) = n\log(1+e^\theta).$$

There is no dispersion parameter, so we will take $\phi \equiv 1$ and $a(\phi) = \phi \equiv 1$.

Poisson distribution: $Y \sim Poisson(\mu)$

$$\begin{aligned} f(Y;\theta,\phi) &= \frac{\mu^Y}{Y!}e^{-\mu} \\ &= \exp\left\{Y \log(\mu) - \mu - \underline{\log(Y!)}\right\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \uparrow C(Y,\phi). \end{aligned}$$

Here

$$\theta = \log(\mu) \Rightarrow \mu = \exp(\theta)$$

and

$$b(\theta) = \mu = \exp(\theta)$$

and there is no additional dispersion parameter. We will take $\phi \equiv 1$ and $a(\phi) = \phi \equiv 1$.

Then,

$$f(Y;\theta,\phi) = \exp\{Y_\theta - \exp(\theta) - \log(\Gamma(Y+1))\}$$

Two parameter gamma distribution: $Y \sim \text{gamma}(\alpha, \beta)$

$$f(Y; \theta, \phi) = \frac{Y^{\alpha-1} \exp(-Y/\beta)}{\beta^\alpha \Gamma(\alpha)}$$

$$= \exp\left\{ \frac{-Y}{\beta} + (\alpha - 1) \log(Y) - \alpha \log(\beta) - \log(\Gamma(\alpha)) \right\}$$

$$= \exp\left\{ \frac{Y\left(\frac{-1}{\alpha\beta}\right) + \log\left(\frac{1}{\alpha}\beta\right)}{\alpha^{-1}} \right.$$

$$\left. \underline{+ \left[(\alpha - 1) \log(Y) + \alpha^{-1} \log(\alpha^{-1}) - \log(\Gamma(\alpha)) \right]} \right\}$$

$$\uparrow C(Y, \phi)$$

Here, $\theta = \frac{-1}{\alpha\beta}$

$$\phi = \alpha^{-1}, \quad b(\theta) = -\log(-\theta), \quad a(\phi) = \phi$$

Inverse Gaussian Distribution: $Y \sim IG(\mu, \sigma^2)$

$$f(Y;\theta,\phi) = \frac{1}{\sqrt{2\pi\sigma^2 Y^3}} \exp\left\{\frac{-(Y-\mu)^2}{2\sigma^2\mu^2 Y}\right\}$$

for $Y > 0, \mu > 0, \sigma^2 > 0$

also $E(Y) = \mu, Var(Y) = \sigma^2\mu^3$

$$\exp\left\{\frac{Y(\frac{-1}{2\mu^2}) - \frac{1}{\mu}}{\sigma^2} + \underline{\left[\frac{1}{2\sigma^2 Y} - \frac{1}{2}\log(2\pi\sigma^2 Y^3)\right]}\right\}$$

$$\uparrow C(Y,\phi)$$

Here,

$$\begin{aligned}
\theta &= \frac{-1}{2\mu^2} \\
b(\theta) &= = \sqrt{-2\theta} \\
\phi &= \sigma^2 \qquad \text{“dispersion parameter”} \\
a(\phi) &= \phi = \sigma^2
\end{aligned}$$

Canonical form of the log-likelihood:

$$\begin{aligned}\ell(\theta, \phi; Y) &= \log(f(Y; \theta, \phi)) \\ &= \frac{Y\theta - b(\theta)}{a(\phi)} + C(Y, \phi)\end{aligned}$$

paratial derivatives:

(i) $\frac{\partial \ell(\theta,\phi;Y)}{\partial \theta} = \frac{Y - b'(\theta)}{a(\phi)}$

(ii) $\frac{\partial^2 \ell(\theta,\phi;Y)}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}$

Apply the following results

$$Ed\left(\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right) = 0$$

$$E\left(\frac{\partial^2 \ell(\theta, \phi; Y)}{\partial \theta^2}\right) + E\left[\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right]^2 = 0$$

to obtain

$$0 = E\left(\frac{\partial \ell(\theta, \phi; Y}{\partial \theta}\right) = \frac{E(Y) - b'(\theta)}{a(\phi)}$$

which implies

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

Furthermore, from (i), (ii) and (iv) we have

$$\begin{aligned}
0 &= E\left(\frac{\partial^2 \ell(\theta, \phi; Y)}{\partial \theta^2}\right) + E\left(\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right)^2 \\
&= \frac{-b''(\theta)}{a(\phi)} + Var\left(\frac{Y - b'(\theta)}{a(\theta)}\right) \\
&= \frac{-b''(\theta)}{a(\theta)} + \frac{Var(Y)}{[a(\phi)]^2}
\end{aligned}$$

which implies

$$V(Y) = \underline{a(\phi)} \qquad \underline{b''(\theta)}$$

$\uparrow$ function of the dispersion parameter

$\uparrow$ called the variance function

Normal distribution: $Y \sim N(\mu, \sigma^2)$

$$\begin{array}{rcl} b(\theta) & = & \frac{\theta^2}{2} \\ b'(\theta) & = & \theta \quad \Rightarrow \quad \mu \equiv E(Y) = \theta \end{array}$$

$$\left.\begin{array}{rcl} b''(\theta) & = & 1 \\ a(\phi) & = & \phi = \sigma^2 \end{array}\right\} \Rightarrow \begin{array}{rcl} \text{Var}(Y) & = & b''(\theta)a(\phi) \\ & = & (1)\phi \\ & = & \phi = \sigma^2 \end{array}$$

$\uparrow$ the variance does not depend on the parameter that define the mean

Binomial distribution: $Y \sim Bin(n, \pi)$

$$\begin{aligned} \theta &= log(\tfrac{\pi}{1-\pi}) \\ b(\theta) &= (n) log(1 + e^{\theta}) \\ b'(\theta) &= \tfrac{ne^{\theta}}{1+e^{\theta}} \Rightarrow E(Y) = \tfrac{ne^{\theta}}{1+e^{\theta}} = n\pi \end{aligned}$$

$$\left.\begin{aligned} b''(\theta) &= \tfrac{ne^{\theta}}{[1+e^{\theta}]^2} \\ a(\phi) &\approx 1 \end{aligned}\right\} \quad \begin{aligned} \text{Var}(Y) &= \tfrac{ne^{\theta}}{[1+e^{\theta}]^2}(1) \\ &= n\left(\tfrac{e^{\theta}}{1+e^{\theta}}\right)\left(\tfrac{1}{1+e^{\theta}}\right) \\ &= n\pi(1-\pi) \end{aligned}$$

$\uparrow \quad Var(Y)$

is completely determined by the parameter that defines the mean.

Poisson Distribution: $Y \sim Poisson(\mu)$

$$\begin{aligned} b(\theta) &= \exp(\theta) \\ b'(\theta) &= \exp(\theta) \Rightarrow \mu = E(Y) = \exp(\theta) \end{aligned}$$

$$\left.\begin{aligned} b''(\theta) &= \exp(\theta) \\ a(\phi) &= \theta \equiv 1 \end{aligned}\right\} \quad \begin{aligned} \text{Var}(Y) &= \exp(\theta) \\ &= E(Y) \\ &= \mu \end{aligned}$$

Gamma Distribution:

$$\begin{aligned} b(\theta) &= -log(-\theta) \\ b'(\theta) &= -\frac{1}{\theta} \Rightarrow E(Y) = -\frac{1}{\theta} \end{aligned}$$

$$\left.\begin{aligned} b''(\theta) &= \tfrac{1}{\theta^2} \\ a(\phi) &= \phi \end{aligned}\right\} \quad \begin{aligned} \text{Var}(Y) &= a(\phi)b''(\theta) \\ &= \tfrac{\phi}{\theta^2} \\ &= \phi[E(Y)]^2 \end{aligned}$$

Canonical link function:
is the link function $h(\mu) = \mathbf{X}^T\boldsymbol{\beta}$ that corresponds to

$$\begin{aligned}\theta &= \mathbf{X}^T\boldsymbol{\beta}\\ \mu &= E(Y) = \frac{\partial b(\theta)}{\partial\theta} = b'(\theta)\end{aligned}$$

<u>Normal (Gaussian) distribution</u>:

$$\begin{aligned}\theta &= \mathbf{X}^T\boldsymbol{\beta}\\ \mu &= b'(\theta) = \theta\end{aligned}$$

then $E(Y) = \mu = \theta = \mathbf{X}^T\beta$
and the canonical link function is the identity function

$$h(\mu) = \mu = \mathbf{X}^T\beta$$

Binomial distribution:

$$\begin{aligned} \theta &= \mathbf{X}^T\boldsymbol{\beta} \\ n\pi &= E(Y) = b'(\theta) = \frac{ne^{\theta}}{1+e^{\theta}} \\ &\Rightarrow \theta = log\Big(\frac{\pi}{1-\pi}\Big) = \mathbf{X}^T\boldsymbol{\beta} \end{aligned}$$

Hence, the canonical (default) link is the logit link function

$$h(\pi) = log\Big(\frac{\pi}{1-\pi}\Big) = \mathbf{X}^T\boldsymbol{\beta}$$

Other available link functions:

Probit: $h(\pi) = \Phi^{-1}(\pi) = \mathbf{X}^T\boldsymbol{\beta}$

where

$$\pi = \Phi(\mathbf{X}^T\boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{X}^T\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}^2} du$$

Complimentary log-log link (clog log):

$$h(\pi) = log[-log(1-\pi)] = \mathbf{X}^T\boldsymbol{\beta}$$

The inverse of this link function is

$$\pi = 1 - \exp[-\exp(\mathbf{X}^T\boldsymbol{\beta})]$$

The glm( ) function

glm(model, family, data, weights, control)

family = binomial
family = binomial(link=logit)
family = binomial(link=probit)
family = binomial(link=cloglog)

Poisson regression (log-linear models)

$$
\begin{aligned}
Y &\sim Poisson(\mu) \\
\theta &= \mathbf{X}^T\boldsymbol{\beta} \\
\mu &= E(Y) = b'(\theta) = \exp(\theta)
\end{aligned}
$$

and the canonical link function is

$$\theta = log(\mu) = \mathbf{X}^T\boldsymbol{\beta}$$

This is called the log link function. Other available link functions:

<u>identity link</u>: $h(\mu) = \mu = \mathbf{X}^T\boldsymbol{\beta}$

<u>square root link</u>: $h(\mu) = \sqrt{\mu} = \mathbf{X}^T\boldsymbol{\beta}$

glm(model, family, data, weights, control)

family = poisson
family = poisson(link=log)
family = poisson(link=identity)
family = poisson(link=sqrt)

Gamma distribution:

$$\begin{aligned}\theta &= \mathbf{X}^T\boldsymbol{\beta} \\ E(Y) &= \mu = \frac{-1}{\theta}\end{aligned}$$

the canonical link function is

$$h(\mu) = \frac{-1}{\mu} = \mathbf{X}^T\boldsymbol{\beta}$$

This is called the "inverse" link.

glm(model, family, data, weights, control)

family = gamma
family = gamma(link=inverse)
family = gamma(link=identity)
family = gamma(link=log)

Inverse Gaussian Distribution

$$\begin{aligned}\theta &= \mathbf{X}^T\boldsymbol{\beta} \\ b(\theta) &= -\sqrt{-2\theta} \\ b'(\theta) &= \frac{1}{\sqrt{-2\theta}} = E(Y) = \mu\end{aligned}$$

The canonical link is

$$\theta = h(\mu) = \frac{-1}{2\mu^2} = \mathbf{X}^T\boldsymbol{\beta}$$

This is the only built-in link function for the inverse gaussian distribution.

glm(model, family, data, weights, controls)

family = inverse.guassian

family = inverse.guassian(link=1/μ^2)

Normal-theory Gauss-Markov model:

$$\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$$

or $Y_1, Y_2, \ldots, Y_n$ are independent with $\mathbf{Y}_j \sim N(\mathbf{X}_j^T \beta, \sigma^2)$. Then

$$f(\mathbf{Y}_j; \boldsymbol{\beta}_j \sigma^2) = \frac{}{\sqrt{2\pi}\sigma^2} \exp\left\{ \frac{-(Y_j - \mathbf{X}_j^T \boldsymbol{\beta})^2}{2\sigma^2} \right\}$$

$$= \exp\left\{ \frac{(Y_j)(\mathbf{X}_j^T \boldsymbol{\beta}) - \frac{(\mathbf{X}_j^T \boldsymbol{\beta})^2}{2}}{\sigma^2} \frac{1}{2} \left[\frac{Y_j^2}{\sigma^2} + log(2\pi\sigma^2) \right] \right\}$$

$$\uparrow C(Y_j, \sigma^2)$$

Here,

$b(\boldsymbol{\beta}) = \frac{(\mathbf{X}_j^T \boldsymbol{\beta})^2}{2}$

$\phi = \sigma^2$ and $a(\phi) = \phi$

$E(\mathbf{Y}_j) = \boldsymbol{\mu}_j = \mathbf{X}_j^T \boldsymbol{\beta}$

The link function is the identity function

$$
\begin{aligned}
h(\mu_j) &= \mu_j \\
&= \mathbf{X}_j^T \boldsymbol{\beta}
\end{aligned}
$$

The class of generalized linear models includes the class of linear models:

(i) Systematic part of the model (identity link function)

$$
\begin{aligned}
\mu_j = E(Y_j) &= \mathbf{X}_j^T \boldsymbol{\beta} \\
&= \beta_1 X_{1j} + \ldots + \beta_p X_{p_j}
\end{aligned}
$$

(ii) Stochastic part of the model

$$Y_j \sim N(\mu_j, \sigma^2)$$

and $Y_1 Y_2 \ldots Y_n$ are independent.

Example 13.1 Mortality of a certain species of beetle after 5 hours exposure to gaseous carbon disulfide (Bliss, 1935).

Dose (mg/l)	Log(Dose) (X_j)	Number killed (Y_j)	Number of survivors ($n_j - Y_j$)	Number exposed n_j
49.057	3.893	6	53	59
52.991	3.970	13	47	60
56.911	4.041	18	44	62
60.842	4.108	28	28	56
64.759	4.171	52	11	63
68.691	4.230	53	6	59
72.611	4.258	61	1	62
76.542	4.338	60	0	60

Logistic Regression:

- $Y_1, Y_2, \ldots, Y_8$ are independent random counts with

$$Y \sim Bin(n_j, \pi_j)$$

- $log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$

Comments:

* Here the logit link function is the canonical link function

* Use of the binomial distribution follows from

(i) Each beetle exposed to log-dose X_j responds independently

(ii) Each beetle exposed to log-dose X_j has probability π_j of dying

<u>Maximum Likelihood Estimation</u>: Joint likelihood for $Y_1, Y_2, \ldots, Y_k$

$$
\begin{aligned}
L(\boldsymbol{\beta}; \mathbf{Y}, X) &= \pi_{j=1}^{k} f(Y_j; \theta_j, \phi) \\
&= \pi_{j=1}^{k} \exp\left\{\frac{Y_j\theta_j - b(\theta_j}{a(phi)} + C(Y_j, \phi)\right\} \\
&= \pi_{j=1}^{k} \exp\Big\{Y_j(\beta_0 + \beta_1 X_j) \\
&\quad -n_j log(1 + e^{\beta_0 + \beta_1 X_j}) + log\binom{n_j}{Y_j}\Big\}
\end{aligned}
$$

Since $\phi = 1$

$a(\phi) = \phi = 1$

$\theta_j = log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$

$b(\theta_j) = n_j log(1 + e^{\beta_0 + \beta_1 X_j})$

$C(Y_j, \phi) = log\binom{n_j}{Y_j}$

log-likelihood function:

$$
\begin{aligned}
\ell(\boldsymbol{\beta}) &= \sum_{j=1}^{k}\left[\frac{Y_j\theta_j - b(\theta_j)}{a(\phi)} + C(Y_j,\phi)\right] \\
&= \sum_{j=1}^{k}\Big[Y_j(\beta_0+\beta_1 X_j) - n_j log(1+e^{\beta_0+\beta_1 X_j}) \\
&\quad + log\binom{n_j}{Y_j}\Big) \\
&= \beta_0\sum_{j=1}^{k} Y_j + \beta_1\sum_{j=1}^{k} Y_j X_j \\
&\quad - \sum_{j=1}^{k} n_j log(1+e^{\beta_0+\beta_1 X_j}) + \sum_{j=1}^{k} log\binom{n_j}{Y_j}
\end{aligned}
$$

To maximize the log-likelihood, solve the system of equations obtained by setting first partial derivatives equal to zero.

Estimating Equations (likelihood equations):

$$
\begin{aligned}
0 &= \frac{\partial \ell(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_0} = \sum_{j=1}^{8} Y_j - \sum_{j=1}^{8} n_j \frac{\exp(\widehat{\beta} + \widehat{\beta}_1 X_1}{1 + \exp(\widehat{\beta} + \widehat{\beta}_1 X_1)} \\
&= \sum_{j=1}^{8} (Y_j - n_j \widehat{\pi}_j) \\
0 &= \frac{\partial \ell(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_0} = \sum_{j=1}^{8} X_j Y_j - \sum_{j=1}^{8} n_j X_j \frac{\exp(\widehat{\beta} + \widehat{\beta}_1 X_1}{1 + \exp(\widehat{\beta} + \widehat{\beta}_1 X_1)} \\
& \sum_{j=1}^{8} X_j (Y_j - n_j \widehat{\pi}_j)
\end{aligned}
$$

- Generally no closed form expression for the solution
- If a solution exists in the interior of the parameter space ($0 < \pi_j < 1$ for all $j = 1, \ldots, k$) then it is <u>unique</u>.

The likelihood equations (or score function) can be expressed as

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \underline{X^T(\mathbf{Y} - \mathbf{m}) = Q}$$

$\uparrow$ score function where

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_k \end{bmatrix} \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \mathbf{m} = \begin{bmatrix} n_1\pi_1 \\ n_2\pi_2 \\ \vdots \\ n_8\pi_8 \end{bmatrix}$$

and from the inverse link function

$$\pi_j = \frac{\exp(\beta_0 + \beta_1 X_j}{1 + \exp(\beta_0 + \beta_1 X_j}$$

We will use $\widehat{Q}, \widehat{\mathbf{m}}$, and $\widehat{\pi}_j$ to indicate when these quantities are evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$.

Second partial derivatives of the log-likelihood function:

$$
\begin{aligned}
H &= \begin{bmatrix} \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0^2} & \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} \\ \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_1^2} \end{bmatrix} \\
&= X^T V X
\end{aligned}
$$

$\nwarrow$ called the Hessian matrix

where

$$
\begin{aligned}
V &= \begin{bmatrix} n_1\pi_1(1-\pi_1) & & \\ & \ddots & \\ & & n_k\pi_k(1-\pi_k) \end{bmatrix} \\
&= Var(\mathbf{Y})
\end{aligned}
$$

We will use $\widehat{H} = X^T \widehat{V} X$ and $\widehat{V}$ to denote H and V evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$.

Fisher Information matrix:

$$
\begin{aligned}
H &= \begin{bmatrix} -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0^2}\right) & -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1}\right) \\ -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1}\right) & -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_1^2}\right) \end{bmatrix} \\
&= E(X^T V X) \\
&= X^T V X = H
\end{aligned}
$$

Since the second partial derivatives do not depend on $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$.

Iterative procedures for solving the likelihood equations:

- Start with some initial values

$$\widehat{\boldsymbol{\beta}}^{(0)} = \begin{bmatrix} \widehat{\beta}_0^{(0)} \\ \widehat{\beta}_1^{(1)} \end{bmatrix} = \begin{bmatrix} \log\left(\frac{P}{1-P}\right) \\ 0 \end{bmatrix}$$

  where

$$P = \frac{\sum_{j=1}^{k} Y_j}{\sum_{j=1}^{k} n_j}$$

- Newton-Raphson algorithm

$$\widehat{\boldsymbol{\beta}}^{(S+1)} = \widehat{\boldsymbol{\beta}}^{(S)} + \underline{\widehat{H}^{-1}\widehat{\mathbf{Q}}}$$

↑ these are evaluated at $\widehat{\boldsymbol{\beta}}^{(S)}$

- Fisher scoring algorithm

$$\widehat{\boldsymbol{\beta}}^{(S+1)} = \widehat{\boldsymbol{\beta}}^{(S)} + \underline{\widehat{I}^{-1}}\widehat{\mathbf{Q}}$$

$\uparrow$ for the logistic regression model with binomial counts $\widehat{I} = \widehat{H}$

<u>Modification</u>:

- Check if the log-likelihood is larger at $\widehat{\boldsymbol{\beta}}^{(S+1)}$ than at $\widehat{\boldsymbol{\beta}}^{(S)}$.

- If it is, go to the next iteration

- If not, invoke a halving step

$$\widehat{\boldsymbol{\beta}}_{new}^{(S+1)} = \frac{1}{2}(\widehat{\boldsymbol{\beta}}^{(S)} + \widehat{\boldsymbol{\beta}}_{old}^{(S+1)})$$

Large sample inference: (see results 9.1-9.3)

From result 9.2,

$$\widehat{\boldsymbol{\beta}} \dot{\sim} N(\boldsymbol{\beta}, (X^T V^{-1} X)^{-1})$$

when $n_1, n_2, \ldots, n_k$ are all large

$$\left[\text{“large” is } \begin{array}{l} n_j \pi_j > 5 \\ n_j(1 - \pi_j) > 5 \end{array} \right]$$

Estimate the covariance matrix as

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}) = (X^T \widehat{V}^{-1} X)^{-1}$$

where $\widehat{V}$ is V evaluated at $\widehat{\boldsymbol{\beta}}$

For the Bliss beetle data:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} -60.717 \\ 14.883 \end{bmatrix}$$

with

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}) = \begin{bmatrix} S^2_{\widehat{\beta}_0} & S_{\widehat{\beta}_0,\widehat{\beta}_1} \\ S_{\widehat{\beta}_0,\widehat{\beta}_1} & S^2_{\widehat{\beta}_1} \end{bmatrix} = \begin{bmatrix} 26.84 & -6.55 \\ -6.55 & 1.60 \end{bmatrix}$$

Wald tests:

Test $H_0 : \beta_0 = 0$ vs. $H_A : \beta_0 \neq 0$

$$X^2 = \left[\frac{\widehat{\beta}_0 - 0}{S_{\widehat{\beta}_0}}\right]^2 = \left[\frac{-60.717}{5.18}\right]^2 = 137.36$$

$$\text{p-value} = Pr\{\chi^2_{(1)} > 197.36\} < .0001$$

Test $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$

$$X^2 = \left[\frac{\widehat{\beta}_1 - 0}{S_{\widehat{\beta}_1}}\right]^2 = \left[\frac{14.883}{1.265}\right]^2 = 138.49$$

$$\text{p-value} = Pr\{\chi^2_{(1)} > 138.49\} < .0001$$

Approximate $(1-\alpha)\times 100\%$ confidence intervals

$$
\begin{aligned}
\widehat{\beta}_0 &\pm Z_{\alpha/2} S_{\widehat{\beta}_0} \\
\widehat{\beta}_1 &\pm Z_{\alpha/2} S_{\widehat{\beta}_1}
\end{aligned}
$$

$\uparrow$ from the standard normal distribution

$Z_{.025} = 1.96$

Likelihood ratio (deviance) tests:

Model A: $log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0$ and $Y_j \sim Bin(n_j, \pi_j)$

Model B: $log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$ and $Y_j \sim Bin(n_j, \pi_j)$

Model C: $log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \alpha_j$ and $Y_j \sim Bin(n_j, \pi_j)$

Model C simply says that $0 < \pi_j < 1$, $j = 1, \ldots, k$

Null: Deviance residual

$$
\begin{aligned}
\text{Deviance}_{\text{Null}} &= -2\Big[\text{log-likelihood}_{\text{Model A}} \\
&\quad --\text{log-likelihood}_{\text{Model C}}\Big] \\
&= 2\Big[\sum_{j=1}^{k} Y_j log(P_j/\bar{P}) \\
&\quad + \sum_{j=1}^{k} (n_j - Y_j) log\Big(\frac{1-P_j}{1-\bar{P}}\Big)\Big] \\
&\quad \text{with } (k-1) d.f.
\end{aligned}
$$

where

$$
\begin{aligned}
P_j &= \frac{Y_j}{n_j}\ j = 1, \ldots, k \\
\bar{P} &= \frac{\Sigma Y_j}{\Sigma n_j} = \frac{\sum_{j=1}^{k} n_j P_j}{\sum_{j=1}^{k} n_j}
\end{aligned}
$$

are the m.l.e.'s of π_j under models C and A, respectively.

LOOSE: deviance residuals

$$
\begin{aligned}
\text{Deviance}_{\text{LOOSE}} &= -2\Big[\text{log-likelihood}_{\text{Model A}} \\
&\quad --\text{log-likelihood}_{\text{Model B}}\Big] \\
&= 2\Big[\sum_{j=1}^{k} Y_j log\Big(\frac{\widehat{\pi}_j}{\bar{P}}\Big) \\
&\quad + \sum_{j=1}^{k} (n_j - Y_j) log\Big(\frac{1-\widehat{\pi}_j}{1-\bar{P}}\Big)\Big] \\
&= 272.7803 \text{ with } (2-1) = 1 d.f.
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\pi}_j &= \frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j)}{1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j} \\
\bar{P} &= \frac{\exp(\widehat{\beta}_0^*}{1 + \exp(\widehat{\beta}_0^*} = \frac{\sum_{j=1}^{k} Y_j}{\sum_{j=1}^{k} n_j}
\end{aligned}
$$

<u>LOOSE: DEV</u> (Lack-of-fit test)

$$H_0 : log\Big(\frac{\pi_j}{1-\pi_j}\Big) = \beta_0 + \beta_1 X_j \quad \text{(model B)}$$

$$H_A : log\Big(\frac{\pi_j}{1-\pi_j}\Big) = \beta_0 + \alpha_j \quad \text{(model C)}$$

$$\begin{aligned} G^2 &= \text{deviance}_{\text{lack-of-fit}} = \text{deviance}_{\text{NULL}} \\ &\quad -\text{deviance}_{\text{LOOSE}} \\ &= 2\Bigg[\sum_{j=1}^{k} Y_j log\Big(\frac{P_j}{\widehat{\pi}_j}\Big) \\ &\quad + \sum_{j=1}^{k}(n_j - Y_j) log\Big(\frac{1-P_j}{1-\widehat{\pi}_j}\Big)\Bigg] \\ &= 11.42 \end{aligned}$$

on

$$\begin{aligned} [(k-1)-1] &= k - 2 d.f. \\ (&= 8-2 = 6 d.f.) \\ &\quad \text{(p-value} = 0.762) \end{aligned}$$

Pearson chi-square test:

$$\begin{aligned}
X^2 &= \sum_{j=1}^{k} \frac{(Y_j - n_j\widehat{\pi}_j)^2}{n_j\widehat{\pi}_j} \\
&\quad + \sum_{j=1}^{k} \frac{(n_j - Y_j - n_j(1-\widehat{\pi}_j))^2}{n_j(1-\widehat{\pi}_j)} \\
&= \sum_{j=1}^{k} n_j \frac{(P_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} + \sum_{j=1}^{k} n_j \frac{(1 - P_j - [1-\widehat{\pi}_j])^2}{1-\widehat{\pi}_j} \\
&= 10.223
\end{aligned}$$

with d.f.$= k - 2 = 6$
p-value = .1155

Estimates of expected counts under Model B.

j	$n_j\widehat{\pi}_j$	$n_j(1-\widehat{\pi}_j)$
1	3.43	55.57
2	9.78	50.22
3	22.66	39.34
4	33.83	22.17
5	50.05	12.95
6	53.27	5.73
6	59.22	2.78
8	58.74	1.26

Cochran's rule:

- All expected counts should be larger than one.
- At least 80% should be larger than 5.

Estimate mortality rates:

$$\widehat{\pi}_j = \frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j)}{1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j}$$

Use the delta method to compute an approximate standard error

$$\begin{aligned}
\frac{\partial \widehat{\pi}_j}{\partial \widehat{\beta}_0} &= \frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j}{[1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j)]^2} \\
&= \widehat{\pi}_j(1 - \widehat{\pi}_j) \\
\frac{\partial \widehat{\pi}_j}{\partial \widehat{\beta}_0} &= \frac{X_j \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j}{[1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_j)]^2} \\
&= X_j \widehat{\pi}_j(1 - \widehat{\pi}_j)
\end{aligned}$$

Define

$$G_j = \left[\begin{array}{cc} \frac{\partial \widehat{\pi}_j}{\partial \widehat{\beta}_0} & \frac{\partial \widehat{\pi}_j}{\partial \widehat{\beta}_1} \end{array} \right] = [\widehat{\pi}_j(1 - \widehat{\pi}_j) X_j \widehat{\pi}_j(1 - \widehat{\pi}_j)]$$

Then, $S^2_{\widehat{\pi}_j} = G_j[\widehat{Var}(\widehat{\boldsymbol{\beta}})]G_j^T$ and an approximate standard error is

$$S_{\widehat{\pi}_j} = \sqrt{G_j[\widehat{Var}(\widehat{\boldsymbol{\beta}})]G_j^T}$$

An approximate $(1 - \alpha) \times 100\%$ confidence interval for π_j is

$$\widehat{\pi}_j \pm Z_{\alpha/2} S_{\widehat{\pi}_j}$$

Estimate the log-dose that provides a certain mortality rate, say π_0.

$$log\Big(\frac{\pi_0}{1-\pi_0}\Big) = \beta_0 + \beta_1 X_0 \Rightarrow$$

$$X_0 = \frac{log\Big(\frac{\pi_0}{1-\pi_0}\Big) - \beta_0}{\beta_1}$$

The m.l.e. for X_0 is

$$\widehat{X}_0 = \frac{log\Big(\frac{\pi_0}{1-\pi_0}\Big)}{\widehat{\beta}_1} - \widehat{\beta}_0$$

Use the delta method to obtain an approximate standard error:

$$\frac{\partial \hat{X}_0}{\partial \hat{\beta}_0} = \frac{-1}{\hat{\beta}_1}$$
$$\frac{\partial \hat{X}_0}{\partial \hat{\beta}_1} = \frac{\left[log\left(\frac{\pi_0}{1-\pi_0}\right) - \hat{\beta}_0\right]}{\hat{\beta}_1^2}$$

Define $\hat{G} = \left[\frac{\partial \hat{X}_0}{\partial \hat{\beta}_0} \frac{\partial \hat{X}_0}{\partial \hat{\beta}_1}\right]$ and compute

$$S^2_{\hat{X}_0} = \hat{G}\left[\widehat{Var}(\hat{\boldsymbol{\beta}})\right]\hat{G}^T$$

The standard error is

$$S_{\hat{X}_0} = \sqrt{\hat{G}\left[\widehat{Var}(\hat{\boldsymbol{\beta}})\right]\hat{G}^T}$$

and an approximate $(1 - \alpha) \times 100\%$ confidence interval for X_0 is

$$\hat{X}_0 \pm Z_{\alpha/2} S_{\hat{X}_0}$$

Suppose you want a 95% confidence interval for the dose that provides a mortality rate of π_0.

$$\text{dose}_0 = \exp(X_0)$$

The m.l.e. for the does is

$$\widehat{\text{dose}}_0 = \exp(\hat{X}_0)$$

Apply the delta method

$$\frac{\partial \exp(\hat{X}_0)}{\partial \hat{X}_0} = \exp(\hat{X}_0)$$

Then

$$\begin{aligned} S_{\widehat{\text{dose}}_0} &= [\exp(\hat{X}_0)]^2 \widehat{Var}(\hat{X}_0) \\ S_{\widehat{\text{dose}}} &= [\exp(\hat{X}_0)] S_{\hat{X}_0} = (\widehat{\text{dose}}_0) S_{\hat{X}_0} \end{aligned}$$

and an approximate $(1-\alpha) \times 100\%$ confidence interval for dose_0 is

$$\exp(\hat{X}_0) \pm Z_{\alpha/2}[\exp(\hat{X}_0)] S_{\hat{X}_0}$$

Fit a probit model to the Bliss beetle data:

$$\Phi^{-1}(\pi_j) = \beta_0 + \beta_1 X_j$$

where

$$\begin{aligned} X_j &= log(\text{dose}) \\ Y_j &= Bin(n_j, \pi_j) \\ & \quad j = 1, 2, \ldots, k \end{aligned}$$

Note that

$$\pi_j = \int_{-\infty}^{\beta_0 + \beta_1 X_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Fit the complimentary log-log model to the Bliss beetle data

$$log[-log(1-\pi_j)] = \beta_0 + \beta_1 X_j$$

where

$$\begin{aligned} X_j &= log(\text{dose}) \\ Y_j &= Bin(n_j, \pi_j) \end{aligned}$$

Note that

$$\pi_j = 1 - \exp[-e^{\beta_0 + \beta_1 X_j}] \; j = 1, 2, \ldots, k$$

Poisson Regression (Log-linear models)[.2in]
Example 13.2:

$$
\begin{aligned}
Y_{jk} &= \text{number of TA98 salmonella colonies on} \\
&\quad \text{the k-th plate with the j-th dosage} \\
&\quad \text{level of quinoline} \\
X_j &= log_{10}(\text{dose of quinoline}(\mu g)) \\
&\quad X_1 = 0 \\
&\quad X_2 = 1.0 \\
&\quad X_3 = 1.5 \\
&\quad X_4 = 2.0 \\
&\quad X_5 = 2.5
\end{aligned}
$$

and $k = 1, 2, 3$ plates at each ??

Independent Poisson counts:

$$Y_{jk} \sim Poisson(m_j) \quad j = 1, \ldots, 5 \quad k = 1, \ldots, 3$$

where

$$\begin{aligned} Pr\{Y_{jk} = y\} &= \frac{m_j^y}{y!} e^{-m_j} \\ & \text{for } y = 0, 1, 2, \ldots \\ E(Y_{jk} &= m_j \\ Var(Y_{jk}) &= m_j \end{aligned}$$

<u>Log-link function</u>:

$$log(m_j = \beta_0 + \beta_1 X_j$$

Joint likelihood function

$$L(\boldsymbol{\beta}) = \pi_{j=1}^{5}\pi_{k=1}^{3}\frac{m_j^{Y_{jk}}}{Y_{jk}}e^{-m_j}$$

Log-likelihood function

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \sum_{j=1}^{5}\sum_{k=1}^{3}[Y_{jk}log(m_j) - m_j - log(Y_{jk}!)] \\
&= \sum_{j=1}^{5}\sum_{k=1}^{3}[Y_{jk}[\beta_0 + \beta_1 X_j] - e^{\beta_0+\beta_1 X_j} \\
&\quad -log(Y_{jk}!)] \\
&= \beta_0 Y_{00} + \beta_1 \sum_{j=1}^{5} X_j Y_{j0} - 3\sum_{j=1}^{5} e^{\beta_0+\beta X_j} \\
&\quad - \sum_{j=1}^{5}\sum_{k=1}^{3} log(Y_{jk}!)
\end{aligned}$$

Likelihood equations:

$$0 = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_0} = Y_{..} - 3 \sum_{j=1}^{5} e^{\beta_0 + \beta_1 X_j}$$

$$0 = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1} = \sum_{j=1}^{5} X_j Y_{j.} - 3 \sum_{j=1}^{5} X_j e^{\beta_0 + \beta_1 X_j}$$

Use the Fisher scoring algorithm:

$$\mathbf{Y} = \begin{bmatrix} Y_1 1 \\ Y_1 2 \\ \vdots \\ Y_{53} \end{bmatrix} X = \begin{bmatrix} 1 & X_1 \\ 1 & X_1 \\ \vdots & \vdots \\ 1 & X_5 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_0} \\ \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1} \end{bmatrix} = X^T(\mathbf{Y} - \mathbf{m})$$

where

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_1 \\ \vdots \\ m_5 \end{bmatrix} = \exp(X\boldsymbol{\beta})$$

Second partial derivatives of $\ell(\boldsymbol{\beta})$

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0^2} = -3 \sum_{j=1}^{5} e^{\beta_0 + \beta_1 X_j} = -\sum_j \sum_k m_j$$

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} = -3 \sum_{j=1}^{5} e^{\beta_0 + \beta_1 X_j} = -\sum_j \sum_k X_j m_j$$

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_1^2} = -3 \sum_{j=1}^{5} X_j^2 e^{\beta_0 + \beta_1 X_j} = -\sum_j \sum_k X_j^2 m_j$$

Fisher information matrix

$$\begin{aligned} H &= E\begin{bmatrix} \frac{-\partial^2 \ell(\beta)}{\partial \beta_0^2} & \frac{-\partial^2 \ell(\boldsymbol{\beta}}{\partial \beta_0 \partial \beta_1} \\ \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_1^2} \end{bmatrix} \\ &= \begin{bmatrix} \sum_j \sum_k m_j & \sum_j \sum_k X_j m_j \\ \sum_j \sum_k X_j m_j & \sum_j \sum_k X_j^2 m_j \end{bmatrix} \\ &= X^T \boldsymbol{\Delta}_{\mathbf{m}} X \end{aligned}$$

where

$$Var(\mathbf{Y}) = \boldsymbol{\Delta}_{\mathbf{m}} = \begin{bmatrix} m_1 & & & & & \\ & m_1 & & & & \\ & & m_1 & & & \\ & & & m_2 & & \\ & & & & \ddots & \\ & & & & & m_5 \end{bmatrix}$$

Iterations:

$$\boldsymbol{\beta}^{(S+1)} = \boldsymbol{\beta}^{(S)} + H^{-1}\mathbf{Q}$$

- Use halving steps as needed

- Starting values: Use the m.l.e.s for the model with $\beta_1 = 0$, i.e.,

$$Y_{ij} \sim Poisson(m)$$

  are i.e.d. with $m = e^{\beta_0}$. Then,

$$\begin{aligned} \widehat{m}^{(0)} &= \bar{Y}_{..} \quad \text{and} \\ \widehat{\beta}_0^{(0)} &= log(\bar{Y}_{..}) \\ \widehat{\beta}_1^{(0)} &= 0 \end{aligned}$$

Large sample normal approximation for the distribution of $\widehat{\boldsymbol{\beta}}$

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, H^{-1})$$

$\nearrow$ estimate this by substituting $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in H to obtain $\widehat{\beta}$

Estimate the mean number of colonies for a specific log_{10}(dose) of quinoline, say X_*

The m.l.e. is

$$\widehat{m}_* = e^{\widehat{\beta}_0 + \widehat{\beta}_1 X_*}$$

Apply the delta method:

$$\begin{aligned} \frac{\partial \widehat{m}_*}{\partial \widehat{\beta}_0} &= e^{\widehat{\beta}_0 + \widehat{\beta}_1 X_*} = \widehat{m}_* \\ \frac{\partial \widehat{m}_*}{\partial \widehat{\beta}_1} &= X_* e^{\widehat{\beta}_0 + \widehat{\beta}_1 X_*} = X_* \widehat{m}_* \end{aligned}$$

Define

$$\widehat{G} = \left[\begin{array}{cc} \frac{\partial \widehat{m}_*}{\partial \widehat{\beta}_0} & \frac{\partial \widehat{m}_*}{\partial \widehat{\beta}_1} \end{array} \right] = [\widehat{m}_* X_* \widehat{m}_*]$$

Then the estimated variance of

$$\widehat{m}_* = e^{\widehat{\beta}_0 + \widehat{\beta}_1 X_*}$$

is

$$\begin{aligned} S^2_{\widehat{m}_*} &= \widehat{G}\widehat{H}^{-1}\widehat{G}^T \\ &= \widehat{G}[X^T \Delta_{\widehat{\mathbf{m}}} X]^{-1} G^T \end{aligned}$$

The standard error is $S_{\widehat{m}_*} = \sqrt{S^2_{\widehat{m}_*}}$ and an approximate $(1-\alpha) \times 100\%$ confidence interval is

$$\widehat{m}_* \pm Z_{\alpha/2} S_{\widehat{m}_*}$$

Four kinds of residuals:

Response residuals:

$$Y_i - m\widehat{u}_i = \left[\frac{Y_i}{n_i} - \widehat{\pi}_i\right]$$

Pearson residuals:

$$\begin{aligned} \frac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{Var}(Y_i)}} &= \frac{Y_i - n_i\widehat{\pi}_i}{\sqrt{n_i\widehat{\pi}_i(1-\widehat{\pi}_i)}} \quad \text{(binomial)} \\ &= \frac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}} \quad \text{(Poisson)} \end{aligned}$$

Working residuals:

$$\frac{Y_i - \widehat{\mu}_i}{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)} = \frac{Y_i - n\widehat{\pi}_i}{n_i\widehat{\pi}_i(1-\widehat{\pi}_i)} \text{(binomial)}$$

$$= \frac{Y_i - \widehat{\mu}_i}{\widehat{\mu}_i} \quad \text{(Poisson)}$$

For the binomial family with logit link

$$\eta_i = log\left(\frac{\pi_i}{1-\pi_i}\right) = log\left(\frac{n_i\pi_i}{n_i(1-\pi_i)}\right) = \mathbf{X}_i^T\boldsymbol{\beta}$$

$$\mu_i = n_i\pi_i = n_i\frac{\exp(\mathbf{X}_i^T\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i^T\boldsymbol{\beta}} = n_i\frac{\exp(\eta_i)}{1+\exp(\eta_i)}$$

$$\frac{\partial \mu_i}{\partial \eta_i} = n_i\frac{\exp(\eta_i)}{[1+\exp(\eta_i)]^2} = n_i\pi_i(1-\pi_i)$$

Deviance residual:

$$e_i = \text{sign}(Y_i - \widehat{\mu}_i)\sqrt{|d_i|}$$

where d_i is the contribution of the i-th case to the deviance

binomial family

$$G^2 = 2\sum_{i=1}^{k}\left[Y_i log\left(\frac{Y_i}{n_i\widehat{\pi}_i}\right) + (n_i - Y_i) log\left(\frac{n_i - Y_i}{n_i(1-\widehat{\pi}_i)}\right)\right]$$

and

$$\begin{aligned} e_i &= \text{sign}(Y_i - n\widehat{\pi}_i) \\ &\times \sqrt{\left|2\left[Y_i log\left(\frac{Y_i}{n_i\widehat{\pi}_i}\right) + (n_i - Y_i) log\left(\frac{n_i - Y_i}{n_i(1-\widehat{\pi}_i)}\right)\right]\right|} \end{aligned}$$

Poisson family

$$
\begin{aligned}
G^2 &= \text{(deviance)} \\
&= 2 \sum_{i=1}^{k} Y_i log \left(\frac{Y_i}{\widehat{\mu}_i} \right)
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\mu}_i &= \exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + ??) \\
log(\mu_i) &= \beta_0 + \beta_1 X_{1i} + \ldots
\end{aligned}
$$

then

$$
e_i = \text{sign}(Y_i - \widehat{\mu}_i) \sqrt{2 Y_i log \left(\frac{Y_i}{\widehat{\mu}_i} \right)}
$$