Introduction to Generalized Linear Models (GLM)

Jaehyoung Hong

Outline

1. Definition of generalized linear model (GLM)
   - Three components
2. GLM for binary response
   - Logistic regression
3. GLM for counts and rates
   - Poisson GLM and negative binomial GLM
4. Likelihood equations and covariance matrix of ML parameter estimates
5. Inference and model checking for GLM
6. Fitting GLM
7. Quasi-likelihood and GLM

Three components of GLM : Random and systemic component and link function

1. Random component : Response variable Y and its probability distribution
2. Systemic component : Explanatory variables
3. Link function : Function of $E(Y)$ that the model equates to the linear predictor

$(y_1, \dots y_N)$: Independent observations from natural exponential family

$$f(y_i; \theta_i) = a(\theta_i) b(y_i) \exp[y_i Q(\theta_i)].$$

<Pdf of Natural exponential family>

$(\eta_1, \dots, \eta_N)$: Linear predictor → Linear combination of explanatory variables

x_{ij} : Value of explanatory variable j ($j = 0,1,2, \dots$) for subject i

$$\eta_i = \sum_j \beta_j x_{ij}, \qquad i = 1, \dots, N.$$

Three components of GLM : Random and systemic component and link function

1. Random component : Response variable Y and its probability distribution
2. Systemic component : Explanatory variables
3. Link function : Function of $E(Y)$ that the model equates to the linear predictor

$(y_1, \ldots y_N)$: Independent observations from natural exponential family

$$f(y_i;\theta_i) = a(\theta_i)b(y_i)\exp[y_i Q(\theta_i)].$$

<Pdf of Natural exponential family>

Let $\mu_i = E(Y_i)$ for $i = 1, \ldots, N$

$$g(\mu_i) = \eta_i = \sum_j \beta_j x_{ij}$$

Link function g is monotonic, differentiable

If $g(\mu_i) = Q(\theta_i)$, then we call g : canonical link

$$\boldsymbol{g(E(Y_i)) = \sum_j \beta_j x_{ij}}$$

<GLM>

Example : Binary logit models for binary data

<Logit models>

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

$\rightarrow \theta = \pi\ , a(\pi) = 1-\pi, b(y) = 1, Q(\pi) = \log[\frac{\pi}{1-\pi}]$

$E(Y) = \pi \rightarrow g(\mu) = \mathrm{Q}(\pi)$: Canonical link

Logistic Regression models or ***Logit model***

Example2 : Poisson loglinear models for count data

<Poisson loglinear models>

$$f(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} = \exp(-\mu)\left(\frac{1}{y!}\right)\exp(y\log\mu), \qquad y = 0, 1, 2, \ldots .$$

$\rightarrow \theta = \mu\,, a(\mu) = \exp(-\mu), b(y) = 1/y!, Q(\mu) = \log(\mu)$

$E(Y) = \mu \rightarrow g(\mu) = \mathrm{Q}(\mu)$: Canonical link

$$\log \mu_i = \sum_j \beta_j x_{ij}, \qquad i = 1, \ldots, N.$$

Poisson loglinear model

Type of GLM for statistical analysis

TABLE 4.1 Types of Generalized Linear Models for Statistical Analysis

Random Component	Link	Systematic Component	Model	Chapters
Normal	Identity	Continuous	Regression	
Normal	Identity	Categorical	Analysis of variance	
Normal	Identity	Mixed	Analysis of covariance	
Binomial	Logit	Mixed	Logistic regression	5 and 6
Poisson	Log	Mixed	Loglinear	8 and 9
Multinomial	Generalized logit	Mixed	Multinomial response	7

Linear probability model : GLM to binary data

Y : $Binary\ response$

$E(Y) = P(Y = 1) = \pi(\boldsymbol{x})$

$\boldsymbol{x} = (x_1, \dots, x_p)$: Explanatory variables

$\Rightarrow$ $var(Y) = \pi(x)[1 - \pi(x)]$

<*Linear probability model*>

$$\pi(x) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

Can you find problem?

($\pi(x) > 1\ or\ \pi(x) < 0$ is possible)

Example : Snoring and he

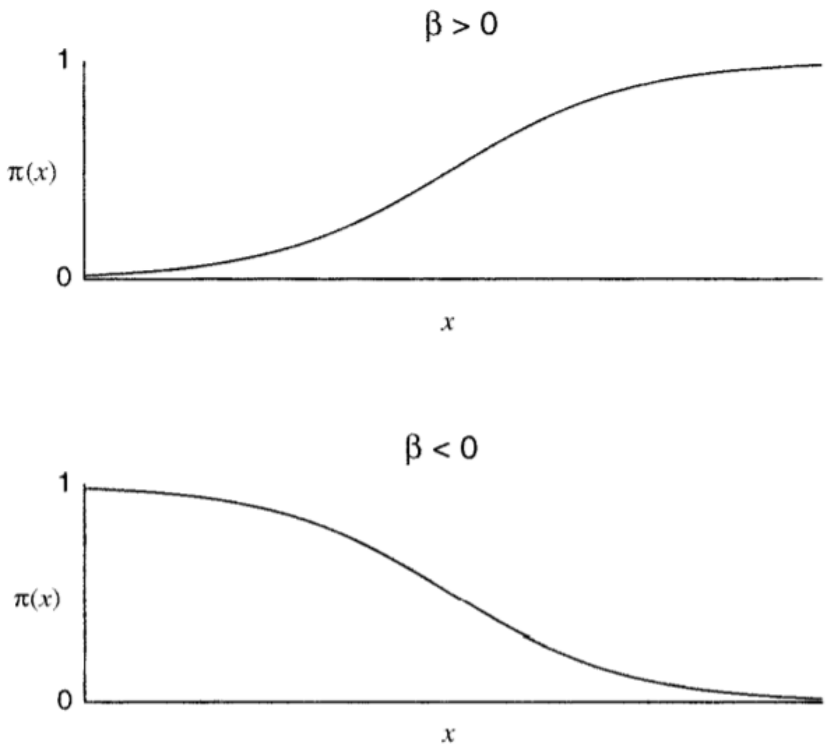

FIGURE 4.2 Logistic regression functions.

TABLE 4.2 Relationship between Snoring and Heart Disease

Snoring	Heart Disease Yes	Heart Disease No	Proportion Yes
Never	24	1355	0.017
Occasionally	35	603	0.055
Nearly every night	21	192	0.099
Every night	30	224	0.118

[a]Model fits refer to proportion of yes responses.

Source: P. G. Norton and E. V. Dunn, *British Med. J.* **291**: 630–63
Group.

<Logistic fit>

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

$(\pi(x) \in [0,1])$

$$\Rightarrow \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

$$\Rightarrow logit[\hat{\pi}(x)] = -3.87 + 0.40x$$

<Linear fit>

Use scores (0, 2, 4, 5) for snoring

$$\Rightarrow \hat{\pi}(x) = 0.0172 + 0.0198x$$

(with $\hat{\beta} = 0.0198$ having $SE = 0.0028$)

Example : Probit and invers cdf Link functions

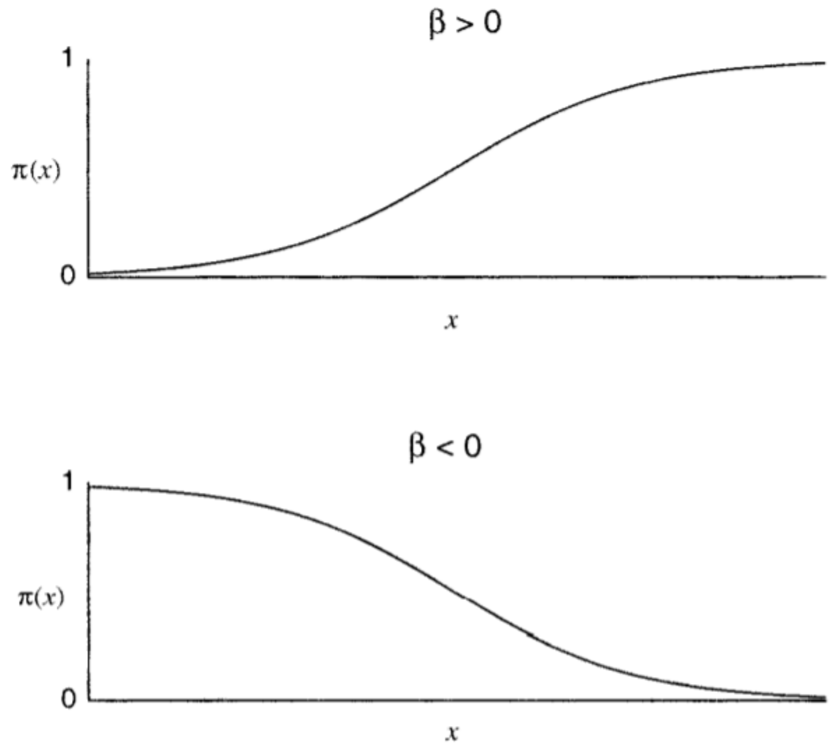

FIGURE 4.2 Logistic regression functions.

<Probit model>

$\Phi(\cdot)$: standard cdf of the class

$$\pi(x) = \Phi(\alpha + \beta x)$$

$$\Rightarrow \quad \Phi^{-1}[\pi(x)] = \alpha + \beta x$$

<Logistic distribution>

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad (\beta > 0)$$

$$\Rightarrow \quad \Phi(x) = \frac{\exp[(x-\mu)/\tau]}{1 + \exp[(x-\mu)/\tau]}$$

Cdf

(mean μ & dispersion $\tau > 0$)

Example : Latent tolerance motivation for binary response models

<Latent tolerance motivation>

x : *Dosage level*

$Y = 1$: *Death* $\Rightarrow$ $Y = 1 \Leftrightarrow T \leq x$

T: *threshold*

Let $F(t) = P(T \leq t)$

$\Rightarrow$ $\pi(x) = P(Y = 1|X = x) = P(T \leq x) = F(x)$

$\Rightarrow$ $\pi(x) = F(x) = \Phi[\frac{x-\mu}{\sigma}]$

$\Rightarrow$ $\Phi^{-1}[\pi(x)] = \alpha + \beta x$

Poisson Loglinear Models

<*Poisson loglinear models*>

$$\log \mu(x) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

$$\Rightarrow \quad \mu(x) = \exp(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p) = e^{\alpha}(e^{\beta_1})^{x_1} \cdots (e^{\beta_p})^{x_p}$$

TABLE 4.3 Number of Crab Satellites by Female's Characteristics[a]

C	S	W	Wt	Sa	C	S	W	Wt	Sa	C	S	W	Wt	Sa	C	S	W	Wt	Sa
2	3	28.3	3.05	8	3	3	22.5	1.55	0	1	1	26.0	2.30	9	3	3	24.8	2.10	0
3	3	26.0	2.60	4	2	3	23.8	2.10	0	3	2	24.7	1.90	0	2	1	23.7	1.95	0
3	3	25.6	2.15	0	3	3	24.3	2.15	0	2	3	25.8	2.65	0	2	3	28.2	3.05	11
4	2	21.0	1.85	0	2	1	26.0	2.30	14	1	1	27.1	2.95	8	2	3	25.2	2.00	1

$$\log \hat{\mu} = \hat{\alpha} + \hat{\beta} x = -3.305 + 0.164x.$$

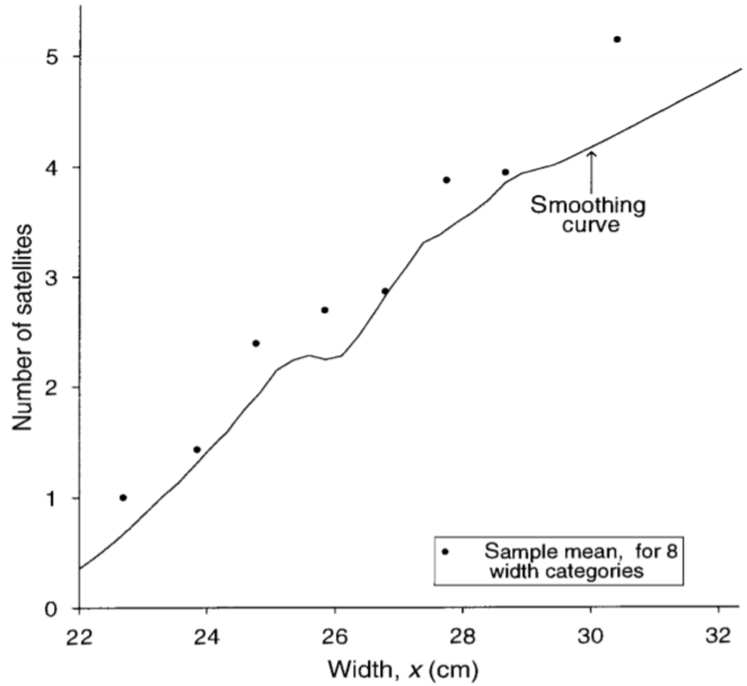

FIGURE 4.4 Smoothings of horseshoe crab counts.

Overdispersion for Poisson GLMs

$<Overdispersion>$

TABLE 4.4 Sample Mean and Variance of Number of Satellites

Width (cm)	Number of Cases	Number of Satellites	Sample Mean	Sample Variance
< 23.25	14	14	1.00	2.77
23.25−24.25	14	20	1.43	8.88
24.25−25.25	28	67	2.39	6.54
25.25−26.25	39	105	2.69	11.38
26.25−27.25	22	63	2.86	6.88
27.25−28.25	24	93	3.87	8.81
28.25−29.25	18	71	3.94	16.88
> 29.25	14	72	5.14	8.29

$Poisson : E = Var$

➡ If true distribution is not poisson, usually variance is larger than mean

➡ 1) Extension of Poisson GLM having extra parameter
2) Quasi-likelihood inference ($Var = \phi\mu$)

Negative binomial GLMs

<Negative binomial GLMs>

$$f(y;k,\mu)=\frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)}\left(\frac{k}{\mu+k}\right)^{k}\left(1-\frac{k}{\mu+k}\right)^{y},\qquad y=0,1,2,\ldots,$$

$k>0, \mu>0$

$\Rightarrow$ $E(Y)=\mu, \quad Var(Y)=\mu+\gamma\mu^2 \quad (\gamma=\frac{1}{k})$

Exponential family (if γ is fixed) + Converge to poisson when $\gamma\to 0$

$Poisson:\ \hat{\mu}=-11.53+0.55x\ (SE=0.06\ for\ \hat{\beta})$

vs

$NB:\ \hat{\mu}=-11.47+0.55x\ (SE=0.12\ for\ \hat{\beta})$

$(\hat{\gamma}=1.07\to var(Y)=\hat{\mu}+\hat{\mu}^2:Geometric\ distribution)$

$\Rightarrow$ NB GLM can capture overdispersion

Poisson regression for rates using offsets

$<GLM\ for\ rates>$

Y_i : $count$ has and index t_i such that its expected value is proportional to t_i

Sample rate : $y_i/t_i \rightarrow E(y_i/t_i) = \mu_i/t_i$

$$\Rightarrow \quad \log(\mu_i/t_i) = \alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

$$\Rightarrow \quad \log(\mu_i)\underbrace{\boxed{-\log(t_i)}}_{offset} = \alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

$$\Rightarrow \quad \mu_i = t_i \exp(\alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})$$

Example : $GLM\ for\ rates$

<$GLM\ for\ rates$>

Table 4.5 Data on Heart Valve Replacement Operations

Age		Type of Heart Valve: Aortic	Type of Heart Valve: Mitral
<55	Number of deaths	4	1
	Time at risk (months)	1259	2082
	Death rate	0.0032	0.0005
55+	Number of deaths	7	9
	Time at risk (months)	1417	1647
	Death rate	0.0049	0.0055

Source: Reprinted with permission, based on data in Laird and Olivier (1981).

$$\log\left({}^{\mu_{ij}}/_{t_{ij}}\right) = \alpha + \beta_1 a_i + \beta_2 v_j, \qquad where\ a_i : age\ indicator\ \&\ v_j : valve\ type\ indicator$$

$$\Rightarrow \quad \widehat{\beta_1} = 1.221\ (SE = 0.514), \qquad \widehat{\beta_2} = -0.330\ (SE = 0.438)$$

Exponential dispersion family

$$<\boldsymbol{Exponential\ dispersion\ family}>$$

$$f(y_i;\theta_i,\phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i,\phi)\}.$$

ϕ : *dispersion parameter,* θ : *natural parameter*

ϕ *is known*

$$f(y_i;\theta_i) = a(\theta_i)b(y_i)\exp[y_iQ(\theta_i)].$$

$$<Natural\ exponential\ family>$$

$$\mathrm{Q}(\theta) = \frac{\theta}{a(\phi)}, \qquad a(\theta) = \exp\left[-\frac{b(\theta)}{a(\phi)}\right], \qquad b(y) = \exp[c(y,\phi)]$$

Usually $a(\phi) = \phi/w_i$ for some known weight w_i

Mean & variance for random component

<**Mean & variance**>

What to do?

$$f(y_i;\theta_i,\phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i,\phi)\}.$$

Express $E(Y_i)$ & $Var(Y_i)$ with above component

➡ $Let\ L_i = \log f(y_i;\theta_i,\phi) \rightarrow L = \sum_i L_i$

➡ $$L_i = [y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i,\phi)\ .$$

$$\partial L_i/\partial\theta_i = [y_i - b'(\theta_i)]/a(\phi), \qquad \partial^2 L_i/\partial\theta_i^2 = -b''(\theta_i)/a(\phi),$$

Mean & variance for random component

<*Mean* & *variance*>

What to do?

$$f(y_i;\theta_i,\phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i,\phi)\}.$$

Express $E(Y_i)$ & $Var(Y_i)$ with above component

$$\partial L_i/\partial\theta_i = [y_i - b'(\theta_i)]/a(\phi), \qquad \partial^2 L_i/\partial\theta_i^2 = -b''(\theta_i)/a(\phi),$$

$$E\left(\frac{\partial L}{\partial\theta}\right) = 0 \quad \text{and} \quad -E\left(\frac{\partial^2 L}{\partial\theta^2}\right) = E\left(\frac{\partial L}{\partial\theta}\right)^2,$$

$$\mu_i = E(Y_i) = b'(\theta_i).$$

$$b''(\theta_i)/a(\phi) = E[(Y_i - b'(\theta_i))/a(\phi)]^2 = \mathrm{var}(Y_i)/[a(\phi)]^2,$$

$$\mathrm{var}(Y_i) = b''(\theta_i)a(\phi).$$

$b(\cdot)$ determines moments of Y_i
(cumulant function)

Mean & variance for Poisson and binomial

$< \boldsymbol{Mean}\ \&\ \boldsymbol{variance\ for\ poisson}\ \&\ \boldsymbol{binomial} >$

$1.\ Y_i\ is\ poisson$

$$f(y_i;\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = \exp(y_i \log \mu_i - \mu_i - \log y_i!)$$
$$= \exp[y_i\theta_i - \exp(\theta_i) - \log y_i!],$$

$< Exponential\ dispersion\ family >$

$$\theta_i = \log(\mu_i),$$
$$b(\theta_i) = \exp(\theta_i),$$
$$a(\phi) = 1,$$
$$c(y_i,\phi) = -\log(y_i!)$$

$$E(Y_i) = b'(\theta_i) = \exp(\theta_i) = \mu_i,$$
$$\operatorname{var}(Y_i) = b''(\theta_i) = \exp(\theta_i) = \mu_i.$$

Mean & variance for Poisson and binomial

$<\boldsymbol{Mean}\ \&\ \boldsymbol{variance\ for\ poisson}\ \&\ \boldsymbol{binomial}>$

2. $n_i Y_i$ *is* $\mathrm{bin}(\mathrm{n_i}, \pi_i)$: y_i is sample proportion (rather than number) of success

$$f(y_i; \pi_i, n_i) = \binom{n_i}{n_i y_i} \pi_i^{n_i y_i} (1 - \pi_i)^{n_i - n_i y_i}$$

$$= \exp\left[\frac{y_i \theta_i - \log[1 + \exp(\theta_i)]}{1/n_i} + \log\binom{n_i}{n_i y_i}\right].$$

$<Exponential\ dispersion\ family>$

$$\theta_i = \log[\pi_i/1 - \pi_i],$$
$$b(\theta_i) = \log[1 + \exp(\theta_i)],$$
$$a(\phi) = 1/n_i,$$
$$c(y_i, \phi) = \log\binom{n_i}{n_i y_i}$$

$$E(Y_i) = b'(\theta_i) = \exp(\theta_i)/[1 + \exp(\theta_i)] = \pi_i,$$

$$\mathrm{var}(Y_i) = b''(\theta_i)a(\phi) = \exp(\theta_i)/\left\{[1 + \exp(\theta_i)]^2 n_i\right\} = \pi_i(1 - \pi_i)/n_i.$$

Systemic component and link function

<*Systemic component and link function*>

Recall

$$\eta_i = \sum_j \beta_j x_{ij}, \qquad i = 1, \ldots, N.$$

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

X **: Model matrix (in GLM)**

➡ $$\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}, \qquad i = 1, \ldots, N.$$

Likelihood equations for a GLM

<Likelihood equation>

For N independent equations

$$L(\boldsymbol{\beta}) = \sum_i L_i = \sum_i \log f(y_i; \theta_i, \phi) = \sum_i \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_i c(y_i, \phi).$$

The notation of $L(\boldsymbol{\beta})$ reflects the dependence of θ on the model parameters $\boldsymbol{\beta}$

➡ $$\partial L(\boldsymbol{\beta}) / \partial \beta_j = \sum_i \partial L_i / \partial \beta_j = 0$$

for all j

➡ $$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Likelihood equations for a GLM

<Likelihood equation>

$$L(\boldsymbol{\beta}) = \sum_i L_i = \sum_i \log f(y_i; \theta_i, \phi) = \sum_i \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_i c(y_i, \phi).$$

$$\Rightarrow \quad \frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

$$\partial L_i / \partial \theta_i = (y_i - \mu_i)/a(\phi), \qquad \partial \mu_i / \partial \theta_i = b''(\theta_i) = \mathrm{var}(Y_i)/a(\phi).$$

$$\partial \eta_i / \partial \beta_j = x_{ij}.$$

$$\Rightarrow \quad \frac{\partial L_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a(\phi)} \frac{a(\phi)}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}$$

Likelihood equations for a GLM

<Likelihood equation>

$$\partial L(\boldsymbol{\beta})/\partial \beta_j = \sum_i \partial L_i/\partial \beta_j = 0$$

$$\Rightarrow \quad \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \qquad j = 0, 1, 2, \cdots$$

Note

Since $\mu_i = g^{-1}(\sum_j \beta_j x_{ij})$, $\boldsymbol{\beta}$ implicitly has influence on likelihood equation.

The key role of the mean-variance relationship

$<Mean - variance\ relationship>$

$$\sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \qquad j = 0, 1, 2, \cdots$$

➡ Likelihood depends on the distribution of Y_i only through $\boldsymbol{\mu_i}$ and $\boldsymbol{var}(\boldsymbol{Y_i})$.

➡ Variance itself depends on the mean

$var(Y_i) = \nu(\mu_i)$

➡ If Y_i has distribution in the natural exponential family, such relationship characterizes the distribution.

$e.g.\ \nu(\mu_i) = \mu_i$ then necessarily Y_i has the Poisson.

Example : Likelihood equations for Binomial GLMs

<$Likelihood\ eq\ for\ bin(n_i, \pi_i)$>

$$\sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 0, 1, 2, \cdots$$

Recall,

$$\pi_i = \Phi\left(\sum_j \beta_j x_{ij} \right),$$

Then, $\pi_i = \mu_i = \Phi(\eta_i) = \sum_j \beta_j x_{ij}$ gives

$$\Rightarrow \quad \partial \mu_i / \partial \eta_i = \phi(\eta_i) = \phi\left(\sum_j \beta_j x_{ij} \right), \qquad \phi(u) = d\Phi(u)/du$$

$$\Rightarrow \quad \sum_i \frac{n_i(y_i - \pi_i)x_{ij}}{\pi_i(1 - \pi_i)} \phi\left(\sum_j \beta_j x_{ij} \right) = 0,$$

since $var(Y_i) = \pi_i(1 - \pi_i)/n_i$

Example : Likelihood equations for Binomial GLMs

<$Likelihood\ eq\ for\ bin(n_i, \pi_i)$>

$$\sum_i \frac{n_i(y_i - \pi_i)x_{ij}}{\pi_i(1-\pi_i)} \phi\left(\sum_j \beta_j x_{ij}\right) = 0,$$

$$\Rightarrow \quad \sum_i n_i(y_i - \pi_i)x_{ij} = 0,$$

When Φ : $standard\ logistic\ cdf$ & $\eta_i = \log[\frac{\pi_i}{1-\pi_i}]$

Asymptotic Covariance matrix

<Covariance matrix of the $\widehat{\boldsymbol{\beta}}$>

Recall, $cov(\widehat{\boldsymbol{\beta}}) = J^{-1}$ $J_{hj} = E[-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_h \partial\beta_j}]$

$\boxed{\text{Fact}}$ $$E\left(\frac{\partial^2 L_i}{\partial\beta_h \partial\beta_j}\right) = -E\left(\frac{\partial L_i}{\partial\beta_h}\right)\left(\frac{\partial L_i}{\partial\beta_j}\right),$$ holds for exponential families

➡
$$E\left(\frac{\partial^2 L_i}{\partial\beta_h \partial\beta_j}\right) = -E\left[\frac{(Y_i - \mu_i)x_{ih}}{\text{var}(Y_i)}\frac{\partial\mu_i}{\partial\eta_i}\frac{(Y_i - \mu_i)x_{ij}}{\text{var}(Y_i)}\frac{\partial\mu_i}{\partial\eta_i}\right]$$
$$= \frac{-x_{ih}x_{ij}}{\text{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2.$$

➡
$$E\left(-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_h \partial\beta_j}\right) = \sum_{i=1}^{N}\frac{x_{ih}x_{ij}}{\text{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2.$$

Asymptotic Covariance matrix

$\boldsymbol{<Covariance\ matrix\ of\ the\ \hat{\beta}>}$

$$E\left(-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_h \partial\beta_j}\right) = \sum_{i=1}^{N} \frac{x_{ih}x_{ij}}{\mathrm{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2.$$

$\boldsymbol{W}$: $diagonal\ matrix$ with $w_i = \left(\partial\mu_i/\partial\eta_j\right)^2/var(Y_i)$

$\Rightarrow$ $$\boldsymbol{J = X^T W X}$$

$\Rightarrow$ $$\widehat{cov}(\hat{\boldsymbol{\beta}}) = \hat{J}^{-1} = \left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}$$

Example : $\widehat{cov}(\widehat{\boldsymbol{\beta}})$ for Poisson loglinear model

<$\widehat{cov}(\widehat{\boldsymbol{\beta}})$ for poisson loglinear>

$$\log \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}.$$

For the log link, $\eta_i = \log(\mu_i)$ & $var(Y_i) = \mu_i$

➡ $$\sum_i (y_i - \mu_i) x_{ij} = 0.$$

These equate the sufficient statistics $\sum_i y_i x_{ij}$ for $\boldsymbol{\beta}$ to their expected values

➡ $$w_i = (\partial\mu_i/\partial\eta_i)^2/\mathrm{var}(Y_i) = \mu_i$$

$\widehat{\boldsymbol{cov}}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X})^{-\mathbf{1}}$ with $\widehat{\boldsymbol{W}}$: $diagonal\ matrix$ with elements of $\widehat{\boldsymbol{\mu}}$ on the main diagonal

Deviance and Goodness of fit

<Deviance and Goodness of fit>

Consider all possible model

$L(\mu; y)$ has its maximum $L(y; y)$: Separate parameter for each observation → *Saturate*

$$\Rightarrow \quad -2\log\frac{\text{maximum likelihood for model}}{\text{maximum likelihood for saturated model}} = -2[\mathrm{L}(\hat{\boldsymbol{\mu}};\mathbf{y}) - \mathrm{L}(\mathbf{y};\mathbf{y})]$$

Let $\tilde{\theta}$ denote the estimate of θ for the saturated model, corresponding to estimates means $\tilde{\mu}_i = y_i$ for all i

$$\Rightarrow \quad \begin{aligned} &-2[L(\hat{\boldsymbol{\mu}};\mathbf{y}) - L(\mathbf{y};\mathbf{y})] \\ &\quad = 2\sum_i \left[y_i\tilde{\theta}_i - b(\tilde{\theta}_i)\right]/a(\phi) - 2\sum_i\left[y_i\hat{\theta}_i - b(\hat{\theta}_i)\right]/a(\phi). \end{aligned}$$

Deviance and Goodness of fit

<Deviance and Goodness of fit>

$$-2[L(\hat{\boldsymbol{\mu}};\mathbf{y}) - L(\mathbf{y};\mathbf{y})]$$
$$= 2\sum_i \left[y_i\tilde{\theta}_i - b(\tilde{\theta}_i)\right]/a(\phi) - 2\sum_i \left[y_i\hat{\theta}_i - b(\hat{\theta}_i)\right]/a(\phi).$$

$a(\phi) = \phi/w_i$

$$\Rightarrow \quad 2\sum_i \omega_i\left[y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right]/\phi = D(\mathbf{y};\hat{\boldsymbol{\mu}})/\phi.$$

We call it, *scaled deviance*

$D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})$: *Deviance*

$\Rightarrow$ Greater scaled deviance, the poorer the fit

Example : Deviance for Poisson GLMs

<Deviance for Poisson GLMs>

Recall, for Poisson GLMs

$\widehat{\theta}_i = \log(\widehat{\mu}_i)$ and $b(\widehat{\theta}_i) = \exp(\widehat{\theta}_i) = \widehat{\mu}_i$ ➡ $\widetilde{\theta}_i = \log(y_i)$ and $b(\widetilde{\theta}_i) = y_i$

➡ $$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_i [y_i \log(y_i/\hat{\mu}_i) - y_i + \hat{\mu}_i].$$

$a(\phi) = 1$

➡ $$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_i y_i \log(y_i/\hat{\mu}_i).$$

$\sum_i y_i = \sum_i \widehat{\mu}_i$

Example : Deviance for binomial GLMs

<Deviance for binomial GLMs>

Recall, for binomial GLMs

$\widehat{\theta}_i = \mathrm{lo}$

With binomial responses, it is possible to construct the data file as expressed here with the counts of successes and failures at each setting for the predictors, or with the individual Bernoulli 0–1 observations at the subject level. The deviance differs in the two cases. In the first case the saturated model has a parameter at each setting for the predictors, whereas in the second case it has a parameter for each subject. We refer to these as *grouped data* and *ungrouped data* cases. The approximate chi-squared distribution for the deviance occurs for grouped data but not for ungrouped data (see Problems 4.22 and 5.37). With grouped data, the sample size increases for a fixed number of settings of the predictors and hence a fixed number of parameters for the saturated model.

$a(\phi) = 1,$

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum \text{observed} \times \log(\text{observed}/\text{fitted}),$$

Likelihood-Ratio Model comparison

$<Likelihood - Ratio\ Model\ Comparison>$

Consider two models, M_0 with fitted values $\hat{\boldsymbol{\mu}}_0$ and M_1 with fitted values $\hat{\boldsymbol{\mu}}_1$, with M_0 a special case of M_1. Model M_0 is said to be *nested* within M_1.

$$-2[L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\hat{\boldsymbol{\mu}}_1;\mathbf{y})]$$

$L(y;y)\ is\ same$

$$= -2[L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\mathbf{y};\mathbf{y})] - \{-2[L(\hat{\boldsymbol{\mu}}_1;\mathbf{y}) - L(\mathbf{y};\mathbf{y})]\}$$

$$= D(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y};\hat{\boldsymbol{\mu}}_1).$$

$a(\phi) = \phi/w_i$

$$D(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y};\hat{\boldsymbol{\mu}}_1) = 2\sum \omega_i\left[y_i\left(\hat{\theta}_{1i} - \hat{\theta}_{0i}\right) - b\left(\hat{\theta}_{1i}\right) + b\left(\hat{\theta}_{0i}\right)\right],$$

Again deviance & approximately a chi-squared (df=diff parameter)

$G^2(M_0|M_1)$

Likelihood-Ratio Model comparison

$<Likelihood - Ratio\ Model\ Comparison>$

Consider two models, M_0 with fitted values $\hat{\boldsymbol{\mu}}_0$ and M_1 with fitted values $\hat{\boldsymbol{\mu}}_1$, with M_0 a special case of M_1. Model M_0 is said to be *nested* within M_1.

$L(y;y)\ is\ same$

$$
\begin{aligned}
&-2[L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\hat{\boldsymbol{\mu}}_1;\mathbf{y})] \\
&\quad = -2[L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\mathbf{y};\mathbf{y})] - \{-2[L(\hat{\boldsymbol{\mu}}_1;\mathbf{y}) - L(\mathbf{y};\mathbf{y})]\} \\
&\quad = D(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y};\hat{\boldsymbol{\mu}}_1).
\end{aligned}
$$

$a(\phi) = \phi/w_i$

$$D(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y};\hat{\boldsymbol{\mu}}_1) = 2\sum \omega_i\left[y_i\left(\hat{\theta}_{1i} - \hat{\theta}_{0i}\right) - b\left(\hat{\theta}_{1i}\right) + b\left(\hat{\theta}_{0i}\right)\right],$$

Again deviance & approximately a chi-squared (df=diff parameter)

$G^2(M_0|M_1)$

Score test for Model comparison

<Score test Model Comparison>

$Var(Y_i) = \nu(\mu_i)$ with $\phi = 1$

$$\Rightarrow \quad X^2 = \sum_i (y_i - \widehat{\mu_i})^2 / \nu(\widehat{\mu_i})$$

$$\Rightarrow \quad X^2(M_0|M_1) = \sum_i \frac{(\widehat{\mu_{1i}} - \widehat{\mu_{0i}})^2}{\nu(\widehat{\mu_{0i}})} \neq X^2(M_0) - X^2(M_1)$$