STAT 8004 Lecture 9

Thursday, April 2, 2015 5:15 PM

- Generalized Linear Models (GLM)

  Model for some more general "response" y variable

  Exponential family for y (univariate random variable)

pdf $$f(y|\theta,\phi) = \exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} - \underbrace{c(y,\phi)}_{\text{can not involve } \theta} \right\}$$

example: normal (μ, σ^2)

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{y(2\mu) - \mu^2/2}{2\sigma^2} - y^2/2\sigma^2\right) \frac{1}{\sqrt{2\pi}\sigma}$$

$$\Rightarrow \theta = \mu, \quad b(\theta) = \mu^2/2 = \theta^2/2 \qquad a(\phi) = \sigma^2, \ \phi = \sigma^2$$

Example $y \sim Bin(n, p)$ with known n

$y \in \{0, 1, 2 \ldots n\}$

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$= \binom{n}{y} \left(\frac{p}{1-p}\right)^y (1-p)^n$$

$$= \exp\left\{ y \log\left(\frac{p}{1-p}\right) + n\log(1-p) + \log\binom{n}{y} \right\}$$

$$\theta = \log\left(\frac{p}{1-p}\right) \quad \Rightarrow \quad p = \frac{\exp(\theta)}{1+\exp(\theta)}$$

$$b(\theta) = -n\log(1-p) = -n\log\left(1 - \frac{\exp(\theta)}{1+\exp(\theta)}\right)$$

$$= n\log(1+\exp(\theta))$$

$$\phi = 1, \quad \ldots \quad c(y,\phi) = \ldots$$

- Example. Poisson $y \in \{0, 1, 2 \ldots\}$

$$f(y|\lambda) = e^{-\lambda} \frac{\lambda^y}{y!}$$

$$= \exp\{ y\log(\lambda) - \lambda - \log(y!) \}$$

$$\theta = \log(\lambda) \qquad \lambda = \exp(\theta)$$

$$\begin{cases} \theta = \log(\lambda) & \lambda = \exp(\theta) \\ b(\theta) = \lambda = \exp(\theta) \\ \phi = 1 \quad , \ \cdots \end{cases}$$

$$\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)}\right\}$$

- Gaussian linear model $\theta = \mu = x^T\beta$ for predictor $x = (x_1, x_2 \cdots x_k)$

- In GLM, 1) model is the exponential family dsn for y

  2) $\underbrace{y_1, y_2 \cdots, y_n}$

  3) $x_1, x_2 \cdots x_n \in \mathbb{R}^k$

To model $\boxed{\theta_i = \boxed{x_i^T\beta}}$

$\{\theta_i\}_{i=1}^n$ $\quad \{x_i^T\beta\}$ $\quad \beta$

example: Binomial $\quad \theta_i = \log\left(\frac{p_i}{1-p_i}\right) = x_i^T\beta$

$$\Rightarrow p_i = \frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}$$

- $\boxed{\text{via } b(\cdot),\ \theta_i \text{ connects to the mean \& variance of } y_i}$

$$f(y|\theta,\phi) \propto \exp\left(\frac{y\theta - b(\theta)}{a(\phi)}\right)$$

$$\log(f(y|\theta,\phi)) = y\theta - b(\theta)/a(\phi) \quad \cdots c(y;\phi)$$

i) $$\frac{\partial \log f(y|\theta,\phi)}{\partial\theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (*)$$

ii) $$\frac{\partial \log f(y|\theta,\phi)}{\partial\theta^2} = \boxed{\frac{-b''(\theta)}{a(\phi)}}$$

$$E_y\left\{\frac{\partial \log f(y|\theta,\phi)}{\partial\theta}\right\} = 0$$

$$\parallel$$

$$\int \frac{\partial \log(f(y))}{\partial\theta} f(y)\,dy$$

$$= \int \frac{1}{f(y)} f'(y)\, f(y)\,dy$$

$$= \int f'(y)\,dy = \left(\int f(y)\,dy\right)' = 0$$

$$E\left(\frac{y - b'(\theta)}{a(\phi)}\right) = 0$$

1) $\underline{\underline{\mu}} = E(y) = \underline{\underline{b'(\theta)}}$

$$\Rightarrow \theta = b'^{-1}(\mu)$$

θ links to μ via $\underbrace{b'^{-1}(\cdot)}_{\text{cannonical link function}}$

$$\underline{\mathrm{var}_y\left(\frac{\partial \log(f(y|\theta,\phi))}{\partial \theta}\right)} = -E\,\frac{\partial^2 \log f(y|\theta,\phi)}{\partial \theta^2}$$

$$\Rightarrow \quad \mathrm{var}\left(\frac{y - b'(\theta)}{a(\phi)}\right) = \frac{b''(\theta)}{a(\phi)}$$

$$\| \qquad\qquad //$$

$$\frac{\mathrm{var}(y)}{a^2(\phi)}$$

(*) $\boxed{\mathrm{var}(y)} = b''(\theta)\,a(\phi)$

recall, $\theta = b'^{-1}(\mu)$ is a function of μ

$$= b''(b'^{-1}(\mu))\,a(\phi)$$

$$\triangleq \boxed{V(\mu)}\,a(\phi)$$

mean-variance relationship.

- $\boxed{\theta_i = x_i^T \beta}$ GLM

- example: Gaussian $b(\theta) = \theta^2/2$ $b''(\theta) = 1$

$$\mathrm{var}(y) = \sigma^2$$

Binomial: $b(\theta) = n \log(1 + \exp(\theta))$

$$b'(\theta) = n\,\frac{\exp(\theta)}{1+\exp(\theta)} \quad (= np)$$

$$b''(\theta) = n\,\frac{\exp(\theta)}{(1+\exp(\theta))^2} \quad (np(1-p))$$

Summary: [illegible]

Summary:

$$\begin{cases} y \sim \text{exponential family} \\ x = (x_1, x_2 \cdots x_k) \quad \text{predictor} \\ (y_1, x_1), (y_2, x_2) \cdots (y_n, x_n) \end{cases}$$

Model: $\theta_i = x_i^T \beta$

Connection $\begin{cases} b'(\theta_i) = \mu_i \\ \theta_i = b'^{-1}(\mu_i) \\ Var(y_i) = b''(\theta_i)\, a(\phi) \end{cases}$

- Objective of GLM: $\begin{cases} \text{- parameter estimation, } \beta, \phi \\ \text{- associated inferences} \\ \text{- predictions} \end{cases}$
  - model building
  - model diagnostic.

- Likelihood framework. (unified for the family of y)

$$Y = (y_1, y_2 \cdots y_n)^T$$

$$X = (X_1, X_2, \cdots X_k) = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{pmatrix}$$

$$f(Y \mid \theta_1, \theta_2 \cdots \theta_n, \phi)$$

$$= \prod_{i=1}^{n} \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} - c(y_i; \phi) \right\}$$

Now: $\theta_i = x_i^T \beta$

$$f(Y \mid \beta, \phi)$$

$$= \prod_{i=1}^{n} \exp\left\{ \frac{y_i x_i^T \beta - b(x_i^T \beta)}{a(\phi)} - \downarrow \right\}$$

$$\log(f(Y|\beta,\phi))$$

$$= \sum_{i=1}^{n} \left(y_i x_i^T\beta - b(x_i^T\beta)\right) / a(\phi) \quad - \downarrow$$

To estimate β

$$\frac{\partial \log(f(Y|\beta,\phi))}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\{y_i x_{ij} - b'(x_i^T\beta)\cdot x_{ij}\}}{a(\phi)}$$

$$= \sum_{i=1}^{n} \frac{1}{a(\phi)} x_{ij} \{ y_i - b'(x_i^T\beta) \}$$

$$0 = \frac{\partial \log f(Y|\beta,\phi)}{\partial \beta} = \sum_{i=1}^{n} x_i \frac{1}{a(\phi)} \{ y_i - b'(x_i^T\beta) \}$$

$\hat{\beta}$ is the root.

- This is a nonlinear function in β generally
- $\hat{\beta}$ is obtained by numerical routines.

- properties of $\hat{\beta}$ (sketch)

$$0 = \sum_{i=1}^{n} x_i \{ y_i - b'(x_i^T\beta) \}$$

- $\hat{\beta}$ is the root of
- $\hat{\beta} \xrightarrow{P} \beta_0$ as $n \to \infty$

$\uparrow$ true

$$= \sum_{i=1}^{n} \underset{p\times 1}{x_i} \{ y_i - b'\{x_i^T(\beta_0 + \hat{\beta} - \beta_0)\} \}$$

$$\simeq \sum_{i=1}^{n} \underset{p\times 1}{x_i} \{ y_i - \left(b'(x_i^T\beta_0) + \underset{1\times p}{\frac{\partial b'(x_i^T\beta)}{\partial \beta^T}} (\hat{\beta} - \beta_0) \right)$$

$$\Rightarrow \sum_{i=1}^{n} x_i \{ y_i - b'(x_i^T\beta_0) \} \simeq \sum_{i=1}^{n} x_i \underset{1\times p}{\frac{\partial b'(x_i^T\beta)}{\partial \beta}} (\hat{\beta} - \beta_0)$$

$$\frac{\partial b'(x_i^T\beta)}{\partial \beta} = b''(x_i^T\beta) \underset{1\times p}{x_i^T}$$

$$\frac{\partial b'(x_i^T\beta)}{\partial \beta} = b''(x_i^T\beta)\, \underset{1\times p}{x_i^T}$$

Now. $\sum_{i=1}^{n} x_i \frac{\partial b'(x_i^T\beta)}{\partial \beta}$

$$= \sum_{i=1}^{n} x_i\, b''(x_i^T\beta)\, x_i^T$$

$$\hat{=} \quad X^T W X$$

$(X = (X_1, X_2 \cdots X_k))$.

$$\underbrace{W = \mathrm{diag}\,(b''(x_1^T\beta), \cdots b''(x_n^T\beta))}$$

$$\sum_{i=1}^{n} x_i (y_i - b'(x_i^T\beta_0)) = X(Y-\mu).$$

$$\Rightarrow \boxed{X^T(Y-\mu) \simeq (X^T W X)(\hat{\beta} - \beta_0)}$$

$$\hat{\beta} - \beta_0 = (X^T W X)^{-1} X^T (Y-\mu)$$

assuming X is of full rank.

$$\mathrm{var}(\hat{\beta} - \beta_0) \approx (X^T W X)^{-1} X^T \underbrace{\mathrm{var}(Y-\mu)}_{W a(\phi)} X (X^T W X)^{-1}$$

$$\mathrm{var}(Y-\mu) = \mathrm{var}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathrm{diag}\,(a(\phi) b''(x_1^T\beta), \cdots a(\phi) b''(x_n^T\beta_0))$$

$$\mathrm{var}(y_i) = b''(\theta_i)\, a(\phi)$$
$$= b''(x_i^T\beta)\, a(\phi)$$

$$= a(\phi)\, W$$

$$\mathrm{var}(\hat{\beta} - \beta_0) = a(\phi)\, (X^T W X)^{-1}$$

Additionally $\hat{\beta} - \beta_0 \overset{AN}{\sim} N(0, a(\phi)(X^T W X)^{-1})$

$\rightarrow W = \mathrm{diag}(b''(x_1^T \underline{\beta_0}), \ldots b''(x_n^T \underline{\beta_0}))$

- Its interval for β?

Estimated variance of $\hat{\beta}$ is

$$a(\hat{\phi})(X^T \hat{W} X)^{-1}$$

$\Rightarrow$ Confidence interval for $c^T\beta$

$$c^T\hat{\beta} \pm z_{\alpha/2} \sqrt{a(\hat{\phi}) c^T (X^T W X)^{-1} c}$$

```
> bliss
  dead alive conc
1    2    28    0
2    8    22    1
3   15    15    2
4   23     7    3
5   27     3    4
```

```
> m1=glm(cbind(dead,alive)~conc,family=binomial,data=bliss)
> summary(m1)

Call:
glm(formula = cbind(dead, alive) ~ conc, family = binomial, data = bliss)

Deviance Residuals:
      1        2        3        4        5
-0.4510   0.3597   0.0000   0.0643  -0.2045

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3238     0.4179  -5.561 2.69e-08 ***
conc          1.1619     0.1814   6.405 1.51e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 64.76327  on 4  degrees of freedom
Residual deviance:  0.37875  on 3  degrees of freedom
AIC: 20.854
```

```
    Null deviance: 64.76327  on 4  degrees of freedom
Residual deviance:  0.37875  on 3  degrees of freedom
AIC: 20.854

Number of Fisher Scoring iterations: 4
```

- $$(X^T W X)\hat{\beta} = X^T(Y-\mu) + (X^T W X)\beta_0$$

$$= X^T W \underbrace{\left(W^{-1}(Y-\mu) + X\beta_0 \right)}_{= Z}$$

$$\hat{\beta}_{new} = \left(X^T W(\hat{\beta}_{old}) X\right)^{-1} X^T W(\hat{\beta}_{old})\, Z(\hat{\beta}_{old})$$

— Doing iterative updating.

— x —

- $\theta_i = x_i^T\beta$, $\boxed{b^{-1\prime}(\mu)} = x_i^T\beta$

- A more broad framework $\Rightarrow \mu = h^{-1}(x_i^T\beta)$

use $\boxed{h(\mu) = x_i^T\beta}$ instead

for Link function $h(\cdot)$.

e.g. $-\log\left(\frac{p}{1-p}\right) = x_i^T\beta$ for Binomial

$$-\Phi^{-1}(p) = x_i^T\beta$$

probit

Φ : cdf of $N(0,1)$

- etc.

- $\underline{\underline{\theta_i}} = b'^{-1}(\mu) = b'^{-1}(h^{-1}(\underline{\underline{x_i^T\beta}}))$

Then, apply the likelihood framework.

- It turns out, the only adjustment needed is one on W_i instead of using $\underline{\underline{b''(\cdot)}}$

```
> m3=glm(cbind(dead,alive)~conc,family=binomial(link=probit),data=bliss)
> summary(m3)

Call:
glm(formula = cbind(dead, alive) ~ conc, family = binomial(link = probit),
    data = bliss)

Deviance Residuals:
       1        2        3        4        5
-0.35863  0.27493  0.01893  0.18230 -0.27545

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.37709    0.22781  -6.045 1.49e-09 ***
conc         0.68638    0.09677   7.093 1.31e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 64.76327  on 4  degrees of freedom
Residual deviance:  0.31367  on 3  degrees of freedom
AIC: 20.789

Number of Fisher Scoring iterations: 4
```

Reference: - Faraway (2006)

- McCullagh & Nelder (1989) GLMs.

- Estimation of ϕ

  - Method of moments

Since $var(y_i) = a(\phi)\,\underline{\underline{b''(x_i^T\beta)}}$

$$\hat{=} \; a(\phi)\, V(\mu_i)$$

$$\mathrm{var}\left(\frac{y_i}{V^{1/2}(\mu_i)} \right) = a(\phi)$$

$$\widehat{a(\phi)} = \frac{1}{n-k} \sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}_i}{V^{1/2}(\hat{\mu}_i)} \right)^2$$

- using likelihood $\ell(\beta, \phi) \leftarrow$ log-likelihood

$$\tilde{\ell}(\phi) = \ell(\hat{\beta}, \phi)$$

Then maximize $\tilde{\ell}(\phi)$ as a function of ϕ.

```
> library(MASS)
> data(epil)
> head(epil)
  y     trt base age V4 subject period      lbase       lage
1 5 placebo   11  31  0       1      1 -0.7563538 0.11420370
2 3 placebo   11  31  0       1      2 -0.7563538 0.11420370
3 3 placebo   11  31  0       1      3 -0.7563538 0.11420370
4 3 placebo   11  31  1       1      4 -0.7563538 0.11420370
5 3 placebo   11  30  0       2      1 -0.7563538 0.08141387
6 5 placebo   11  30  0       2      2 -0.7563538 0.08141387
```

```
> m2=glm(y~lbase*trt+lage+V4,family=poisson,data=epil)
> summary(m2)

Call:
glm(formula = y ~ lbase * trt + lage + V4, family = poisson,
    data = epil)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0915  -1.4126  -0.2739   0.7580  10.7711

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.89791    0.04260  44.552  < 2e-16 ***
lbase               0.94862    0.04360  21.759  < 2e-16 ***
trtprogabide       -0.34588    0.06100  -5.670 1.42e-08 ***
```

```
(Intercept)          1.89791    0.04260  44.552  < 2e-16 ***
lbase                0.94862    0.04360  21.759  < 2e-16 ***
trtprogabide        -0.34588    0.06100  -5.670 1.42e-08 ***
lage                 0.88760    0.11650   7.619 2.56e-14 ***
V4                  -0.15977    0.05458  -2.927  0.00342 **
lbase:trtprogabide   0.56154    0.06352   8.841  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
```

- Diagnostic.

$$\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)}\right\}$$

- Deviance.

We can write the log likelihood as a function of

$\mu = (\mu_1, \mu_2, \dots \mu_n)$

$\rightarrow \underbrace{\ell(Y; \mu, \phi)}$ $\leftarrow$ log likelihood

- A <u>full</u> or satured model is one with one parameter for each observation.

It turns out

$\ell(Y; \underline{\underline{Y}}, \phi)$ is the maximum likelihood of the satured model for given ϕ.

Then with a model $\theta_i = x_i^T \underline{\beta}$ giving me

$\hat{\mu}_i$ with $\underline{\hat{\theta}_i = x_i^T \hat{\beta}}$.

$\hat{\mu}_i$ with $\theta_i = x_i'\beta$.

$$\boxed{2\{\ell(Y; Y, \underline{\underline{\phi}}) - \ell(Y; \hat{\mu}_i, \phi)\}}$$

scaled deviance in GLM

$$= 2\sum_{i=1}^{n} \frac{\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - (b(\tilde{\theta}_i) - b(\hat{\theta}_i))\}}{a(\phi)}$$

- $\hat{\theta}_i$ is the estimation from the model $\theta_i = x_i'\beta$

- $\tilde{\theta}_i$ is the estimation from the saturated model

$$\hat{\mu}_i = y_i, \quad \tilde{\theta}_i = b'^{-1}(y_i)$$

- By likelihood theory

  scaled deviance $\sim \chi^2_{d.f.}$

- $\boxed{\text{Null model}}$ $\hat{\theta}_i = \hat{\beta}_0$

- Since the scaled deviance involving ϕ,

  the computable part is called deviance.

Normal	$\sum(y-\hat{\mu})^2,$
Poisson	$2\sum\{y\log(y/\hat{\mu})-(y-\hat{\mu})\},$
binomial	$2\sum\{y\log(y/\hat{\mu})+(m-y)\log[(m-y)/(m-\hat{\mu})]\},$
gamma	$2\sum\{-\log(y/\hat{\mu})+(y-\hat{\mu})/\hat{\mu}\},$
inverse Gaussian	$\sum(y-\hat{\mu})^2/(\hat{\mu}^2 y).$

```
> m1=glm(cbind(dead,alive)~conc,family=binomial,data=bliss)
> summary(m1)

Call:
glm(formula = cbind(dead, alive) ~ conc, family = binomial, data = bliss)

Deviance Residuals:
      1        2        3        4        5
-0.4510   0.3597   0.0000   0.0643  -0.2045

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3238     0.4179  -5.561 2.69e-08 ***
conc          1.1619     0.1814   6.405 1.51e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 64.76327  on 4  degrees of freedom
Residual deviance:  0.37875  on 3  degrees of freedom
AIC: 20.854

Number of Fisher Scoring iterations: 4
```

```
> m1=glm(cbind(dead,alive)~conc,family=binomial,data=bliss)
> summary(m1)

Call:
glm(formula = cbind(dead, alive) ~ conc, family = binomial, data = bliss)

Deviance Residuals:
      1        2        3        4        5
-0.4510   0.3597   0.0000   0.0643  -0.2045

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3238     0.4179  -5.561 2.69e-08 ***
conc          1.1619     0.1814   6.405 1.51e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 64.76327  on 4  degrees of freedom
Residual deviance:  0.37875  on 3  degrees of freedom
AIC: 20.854

Number of Fisher Scoring iterations: 4

> 1-pchisq(deviance(m1),df.residual(m1))
[1] 0.9445968
```

$2\{$ Scaled deviance $(m1)$ — Scaled deviance $(m2)\}$ $\sim \chi^2_{df(m_1)-df(m_2)}$

m_2 is nested in m_1

- anova (m_1, m_2)

— x —

- Faraway (2006), with estimated ϕ,

— x —

- Residuals:
  - Pearson residuals

$(y_i - \hat{\mu}_i)/\ldots$

- $r_{P,i} = (y_i - \hat{\mu}_i) / V^{\frac{1}{2}}(\hat{\mu}_i)$

- $\sum r_{P,i}^2 \sim \chi^2 \leftarrow$ pearson's χ^2 test statistic.

- Deviance residuals

let D be the <u>deviance</u> $\quad D = \boxed{\sum_{i=1}^{n} d_i^2}$

$r_{D,i} = \text{sign}(y_i - \hat{\mu}_i)\sqrt{d_i} \leftarrow$

$\sum_{i=1}^{n} r_{D,i}^2 = D$

- Just used as <u>residuals</u>.

```
## S3 method for class 'glm'
residuals(object, type = c("deviance", "pearson", "working",
                           "response", "partial"), ...)
```

— x —

Summary : $\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)}\right\}$

$b'(\theta) = \mu = E(y)$

$$\begin{cases} \begin{cases} b'(\theta) = \mu = E(y) \\ b''(\theta)\, a(\phi) = Var(y) \end{cases} \\ \text{MLE} \end{cases}$$

— x —

- Office hours
  - Monday April 6 (1:00 - 2:15pm)
    5:00 - 6:30 pm)
  - Tuesday April 7 (11:00^{am} - 2:00pm).