(10+)

16. Oct. 01

Fisher method of scoring.

L : likelihood

$$\underset{p\times 1}{l} = \underset{p\times 1}{\frac{\partial \log L}{\partial \beta}} \simeq \underset{p\times 1}{\frac{\partial \log L}{\partial \beta_-}} + \underset{p\times p}{\frac{\partial^2 \log L}{\partial \beta \, \partial \beta_-}} \underset{p\times 1}{(\beta - \beta_-)}$$

$$\simeq \frac{\partial \log L}{\partial \beta_-} - I(\beta_-)(\beta - \beta_-)$$

$I(\beta)$: Fisher information

$$\frac{\partial \log L}{\partial \hat{\beta}} = \underset{\sim}{0}$$

$$\beta - \beta_- = I(\beta_-)^{-1} \frac{\partial \log L}{\partial \beta_-}$$

16 Oct 01

For canonical links

$$s(\underset{\sim}{\beta}) = \frac{1}{\phi} \underset{\sim}{X}^T \underset{\sim}{\Omega} [\underset{\sim}{y} - \underset{\sim}{\mu}(\underset{\sim}{\beta})]$$

$$F(\underset{\sim}{\beta}) = \frac{1}{\phi} \underset{\sim}{X}^T \underset{\sim}{\Omega} \underset{\sim}{V}(\underset{\sim}{\beta}) \underset{\sim}{X}$$

$$\underset{\sim}{\Omega} = \text{diag}(w_i), \quad \underset{\sim}{V}(\underset{\sim}{\beta}) = \text{diag}\{v(\mu_i)\}$$

Compute $\hat{\underset{\sim}{\beta}}$ as solution of

$$\underset{\sim}{s}(\underset{\sim}{\beta}) = 0$$

Fisher scoring

$$\hat{\underset{\sim}{\beta}}^{(k+1)} = \hat{\underset{\sim}{\beta}}^{(k)} + \underset{\sim}{F}^{-1}(\hat{\underset{\sim}{\beta}}^{(k)}) \underset{\sim}{s}(\hat{\underset{\sim}{\beta}}^{(k)})$$

If one defines the "working observation vector"

$$\tilde{\underset{\sim}{y}}(\underset{\sim}{\beta}) = \begin{bmatrix} \tilde{y}_1(\beta) \\ \vdots \\ \vdots \end{bmatrix}$$

$$\tilde{y}_i(\beta) = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + D_i^{-1}(\beta)[y_i - \mu_i(\beta)]$$

$$\underset{\sim}{W} =$$

16 Oct 01

then Fisher scoring becomes

$$\hat{\beta}^{(k+1)} = (X^T W^{(k)} X)^{-1} X^T W^{(k)} \tilde{y}^{(k)}$$

i.e. iteratively reweighted least squares.

Regularity assumptions
Uniqueness and existence of mle's
Asymptotic properties

$$\hat{\beta} \sim N(\beta, F^{-1}(\hat{\beta}))$$

$$F^{-1} X^T W \left(X^T \beta + D^{-1} [y - \mu]\right)$$

$$\beta + F^{-1} \underline{X^T W D^{-1} [y - \mu]}$$

$$V = D^2 \sigma^{-2}$$

$$X^T D \sigma^{-2} [y - \mu]$$

18 March 04

glm fitting.

MLE for exponential family may be carried out by IRLS.

The notation,

$$E\{Y|x\} = \mu = h(\eta) \qquad \text{inverse link}$$

$$\eta = \underset{\sim}{x}' \underset{\sim}{\beta}$$

$$= g(\mu) \qquad \text{link}$$

$$\text{Var}\{Y|x\} = v(\mu) \qquad \text{(perhaps added scale)}$$

Write

$$g(Y) \simeq g(\mu) + g'(\mu)(Y-\mu)$$

$$= \underset{\sim}{x}' \underset{\sim}{\beta} + \frac{\partial \eta}{\partial \mu}(Y-\mu)$$

18 March 04

$$E\left\{ \underset{\sim}{x}'\underset{\sim}{\beta} + \frac{\partial \eta}{\partial \mu}(Y-\mu) \right\} \simeq \underset{\sim}{x}'\underset{\sim}{\beta}$$

$$\mathrm{Var}\left\{ \underset{\sim}{x}'\underset{\sim}{\beta} + \frac{\partial \eta}{\partial \mu}(Y-\mu) \right\} \simeq \left(\frac{\partial \eta}{\partial \mu}\right)^2 v(\mu)$$

Regress $\hat{\eta} + (Y-\hat{\mu})\dfrac{\partial \hat{\eta}}{\partial \mu}$ on $\underset{\sim}{x}$

with weights $\left(\dfrac{\partial \mu}{\partial \hat{\eta}}\right)^2 \dfrac{1}{v(\hat{\mu})}$

The quantities with hats are evaluated ~~from~~ the preceding iteration.

One starts with $\hat{\mu}_0 = Y$

21 Oct 01

<u>Summary of glm.</u>

$\eta = \underset{\sim}{x}' \underset{\sim}{\beta}$

$EY = \mu$

$\eta = g(\mu)$ g: <u>link function</u>

Exponential family

φ: dispersion parameter

$(y_i, \underset{\sim}{x}_i)$ $i = 1, \dots, n$

$\text{var}\, Y = \varphi\, v(\mu)$

Particular cases of the glm 21 Oct. 01

1. Bernoulli, Binomial $Y = 0, 1$
$= 0, 1, \ldots, n$ Really Y/n

Natural link $g(\pi) = \log(\pi/1-\pi) = \eta$ logit

but remember $g(\pi) = \Phi^{-1}(\pi) = \eta$ probit

$g(\pi) = \log(-\log(1-\pi))$ clog-log

2. Poisson $\dfrac{\mu^y e^{-\mu}}{y!}$ $Y = 0, 1, 2, \ldots$

Natural link $g(\mu) = \log \mu$
$g(\mu) = \mu$ identity

3. Normal $-\infty < Y < \infty$

Natural link $g(\mu) = \mu = \eta$

$\mu = \eta = x' \beta$

Gauss-Markov + normality

Dispersion parameter $\varphi = \sigma^2$

expresses uncertainty

21. Oct 01

<u>Gamma</u>

$$f(y \mid \mu, \nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1} \exp\left\{-\frac{\nu}{\mu} y\right\} \quad y > 0$$

$$EY = \mu$$

$$\operatorname{var} Y = \mu^2/\nu$$

$$\varphi = 1/\nu$$

Natural link $g(\mu) = 1/\mu = \eta$

$$EY = 1/\underset{\sim}{x}'\underset{\sim}{\beta} \qquad \text{need} > 0$$

Sometimes use

$$g(\mu) = \mu = \eta$$

21 Oct 01

Inverse Gaussian

$$\sqrt{\frac{\sigma^2}{2\pi y^3}} \exp\left\{ - \frac{\sigma^2 (y-\mu)^2}{2\mu^2 y} \right\} \quad y > 0$$

lifetimes

$E\,Y = \mu$

$\mathrm{Var}\,Y = \mu^3 \sigma^2$

Natural link $g(\mu) = 1/\mu^2 = \eta$

$E\,Y = 1 / (\underset{\sim}{x}' \underset{\sim}{\beta})^2$

$\varphi = \sigma^2$

22 Oct 01.

Selecting and Checking Models.

Goodness of fit statistics.

Pearson statistic.

$$X^2 = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 / \mathrm{var}(\hat{\mu}_i)$$

is approx $\varphi \chi^2_{n-p}$ p: no. of estimated params

Deviance or likelihood ratio statistic.

$$D = -2\varphi \sum_{i=1}^{n} \{ l_i(\hat{\mu}_i) - l_i(y_i) \}$$

is approx $\varphi \chi^2_{n-p}$

Problems with approx if eg. $y_i = 1, 0$

Selecting & Checking Models

22 Oct '01

Residuals

$$\text{Deviance} = \sum_{i=1}^{n} \chi^2_D(y_i, \hat{\mu}_i)$$

$$\chi^2_D(y_i, \hat{\mu}_i) = 2\left(l_i(y_i) - l_i(\hat{\mu}_i)\right)$$

$$l_i(y_i) = [\, y_i \theta_i - b(\theta_i)]/\phi + c(y_i, \phi)$$

$$\hat{\mu}_i = \mu_i(\hat{\beta})$$

Deviance residual

$$r_i^D = \operatorname{sgn}\{y_i - \hat{\mu}_i\}\sqrt{\chi^2_D(y_i, \hat{\mu}_i)}$$

Pearson residual

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\operatorname{var}}\, y_i}}$$

often highly skewed

need adjustments: Anscombe, ...

11 22 Oct 01

Variable selection.

Sometimes $\exists$ many potential covariates

Variable selection methods aim at determining submodels with a moderate number of parameters that still fit the data adequately.

Likelihood

$$l(\beta;\theta) = \sum_{i=1}^{n} l_i(x_i;\theta)$$

where θ contains other parameters, e.g. ϕ

$\beta = (\beta_1, \beta_2)$ $\quad H_0: \beta_2 = 0$

All subsets selection:

Akaike Information Criteria.

$$AIC = -2\, l(\tilde{\beta}_1, 0, \tilde{\theta}) + 2(r+s)$$

$r = \dim(\tilde{\beta}_1)$

$s = \dim(\tilde{\theta})$

$\tilde{\beta}_1, \tilde{\theta}$ mle under H_0

22 Oct 06

<u>Stepwise backward,</u>

<u>Stepwise forward.</u>

E

24 Oct 01

Goodness of fit.

Deviances, $D(\underset{\sim}{y};\hat{\mu})$

Normal $\quad \Sigma(y-\hat{\mu})^2$

Poisson $\quad 2\Sigma\{y \log y/\hat{\mu} - (y-\hat{\mu})\}$

Binomial $\quad 2\Sigma\{y \log \frac{y}{\hat{\mu}} + (m-y)\log\frac{m-y}{m-\hat{\mu}}\}$

Gamma $\quad 2\Sigma\{-\log\frac{y}{\hat{\mu}} + \frac{(y-\hat{\mu})}{\hat{\mu}}\}$

Inverse Gaussian $\quad \Sigma(y-\hat{\mu})^2/\hat{\mu}^2 y$

Scaled deviance $\quad D^*(\underset{\sim}{y};\underset{\sim}{\hat{\mu}}) = D(y;\hat{\mu})/\varphi$

Analysis of Deviance ANODEV

$$\frac{D_0 - D_1}{\hat{\varphi}(p-q)} \sim F_{p-q,\,m-p}$$, i.e. Gaussian result

May need a better approximation.

24 Oct. 01

Contingency table data.

Eg. Lung cancer data

Question(s). Dependence of death rate on amount of smoking and number of years smoking.

Eg. Lizard data

Question. Categories independent?

Table 1

Man-years at risk, number of cases of lung cancer (in parentheses), and fitted values obtained under the product model

Years of smoking (age minus 20 years)	Cigarettes/day: Nonsmokers 0	1–9 5.2	10–14 11.2	15–19 15.9	20–24 20.4	25–34 27.4	35 + 40.8	Age fit* (per 100 000 man years)
15–19	10366 (1)	3121	3577	4317	5683	3042	670	.3
20–24	8162	2937	3286 (1)	4214	6385 (1)	4050 (1)	1166	.9
25–29	5969	2288	2546 (1)	3185	5483 (1)	4290 (4)	1482	1.9
30–34	4496	2015	2219 (2)	2560 (4)	4687 (6)	4268 (9)	1580 (4)	8.5
35–39	3512	1648 (1)	1826	1893	3646 (5)	3529 (9)	1336 (6)	8.8
40–44	2201	1310 (2)	1386 (1)	1334 (2)	2411 (12)	2424 (11)	924 (10)	23.2
45–49	1421	927	988 (2)	849 (2)	1567 (9)	1409 (10)	556 (7)	29.4
50–54	1121	710 (3)	684 (4)	470 (2)	857 (7)	663 (5)	255 (4)	46.5
55–59	826 (2)	606	449 (3)	280 (5)	416 (7)	284 (3)	104 (1)	77.3
Smoking effect†	1.0	3.39	8.16	10.1	18.2	22.6	36.8	

* Age fit = $\exp(\hat{\mu} + \hat{\alpha}_j)$, where $\hat{\mu}$ and $\hat{\alpha}$ are ML estimates defined by the product model.
† Smoking effect = $\exp(\hat{\delta}_k)$, where $\hat{\delta}$ is an ML estimate defined by the product model.
The estimated lung cancer deaths per 100 000 man-years in Row j and Column k are given by Fit = Age fit × Smoking effect = $\exp(\hat{\mu} + \hat{\alpha}_j + \hat{\delta}_k)$.

Table 2-2

Counts for Structural Habitat Categories for *sagrei* Adult Male *Anolis* Lizards of Bimini (Schoener [1968])

(a) Observed values

		Perch Diameter (inches)		
		≤ 4.0	> 4.0	Totals
Perch Height (feet)	> 4.75	32	11	43
	≤ 4.75	86	35	121
	Totals	118	46	164

24 Oct. 01

There are 3 distribution commonly used:

1. Poisson

2. Multinomial

3. Product multinomial

In the case of 2 and 3 can use family = poisson