Fisher method of scoring.

\[ L = \frac{\partial \log L}{\partial \beta} = \frac{\partial \log L}{\partial \beta_-} + \frac{\partial^2 \log L}{\partial \beta_- \partial \beta_-} (\beta - \beta_-) \]

\[ = \frac{\partial \log L}{\partial \beta_-} - I(\beta_-) (\beta - \beta_-) \]

\[ I(\beta_-) : \text{Fisher information} \]

\[ \frac{\partial \log L}{\partial \beta_-} = 0 \]

\[ \beta - \beta_- = I(\beta_-)^{-1} \frac{\partial \log L}{\partial \beta_-} \]
For canonical links

\[
\begin{align*}
\Delta(\beta) &= \frac{1}{p} \sum \mathbf{z}_i \mathbf{z}_i \mathbf{y}_i - \mu(\beta) \\
F(\beta) &= \frac{1}{p} \sum \mathbf{z}_i \mu(\beta) \mathbf{z}_i \\
-\mathbf{L} &= \text{diag}(w_i), \quad \mathbf{V}(\beta) = \text{diag}\{v(\mu_i)\}
\end{align*}
\]

Can take \(\hat{\beta}\) as solution of

\[
\frac{\partial L}{\partial \beta} = 0
\]

**Fisher scoring**

\[
\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \mathbf{F}^{-1}(\hat{\beta}^{(k)}) \frac{\partial L}{\partial \beta}^{(k)}
\]

If one defines the "working observation vector"

\[
\tilde{y}_i(\beta) = \begin{bmatrix} \tilde{y}_i(\beta) \end{bmatrix}
\]

\[
\tilde{y}_i(\beta) = \mathbf{x}_i^\top \hat{\beta} + \mathbf{D}_i^{-1}(\beta) \left[ y_i - \mu_i(\beta) \right]
\]
Then Fisher scoring becomes
\[
\hat{\beta}^{(k+1)} = (X^T W^{(k)} X)^{-1} X^T W^{(k)} \tilde{y}^{(k)}
\]
in iteratively reweighted least squares.

Regularity assumptions
Uniqueness and existence of mle's
Asymptotic properties
\[
\hat{\beta} \sim N(\beta, \mathbf{V}^{-1}(\beta))
\]
Glor-fitting.

MLE for exponential family may be carried out by IRLS.

The notation,

\[ E[Y | x] = \mu = h(\eta) \quad \text{inverse link} \]
\[ \eta = x' \beta \]
\[ = g(\mu) \quad \text{link} \]
\[ \text{Var}[Y | x] = \sigma^2 (\mu) \quad (\text{perhaps added scale}) \]

Write

\[ g(Y) = g(\mu) + g'(\mu)(Y - \mu) \]
\[ = x' \beta + \frac{\partial g}{\partial \mu} (Y - \mu) \]
\[ Ef \left( x' \beta + \frac{\partial \hat{y}}{\partial \mu} (1-\mu)^2 \right) = x' \beta \]

\[ \text{Var} \left( \frac{x' \beta + \frac{\partial \hat{y}}{\partial \mu} (1-\mu)^2}{\hat{y}} \right) = \left( \frac{\partial \hat{y}}{\partial \mu} \right)^2 \text{Var}(\mu) \]

\[ \text{Regression} \hat{y} + (1-\hat{\mu}) \frac{\partial \hat{y}}{\partial \mu} \text{ on } x \]

with weights \( \left( \frac{\partial \hat{y}}{\partial \mu} \right)^2 \text{Var}(\mu) \)

The quantities with hats are evaluated from the preceding iteration.

One starts with \( \hat{\mu}_0 = 1 \)
Summary of glm

\[ Y = \beta 'X \]

\[ EY = \mu \]

\[ Y = g(\mu) \quad g: \text{link function} \]

Exponential family

\[ \phi: \text{dispersion parameter} \]

\[ (y_i, x_i) \quad i = 1, \ldots, n \]

\[ \text{var} Y = \phi \nu(\mu) \]
Particular cases of the glm. 21 Oct. 01

1. Bernoulli, Binomial \( y = 0, 1 \)
   \[ y_i = 0, 1, \ldots, n \]
   Really \( Y/n \)
   Natural link \( g(\pi) = \log \left( \frac{\pi}{1-\pi} \right) = \eta \) \text{ logit }
   
   but remember \( g(\pi) = \frac{1}{g'(\pi)} = \eta \) \text{ probit }

\[ g(\pi) = \log \left( -\log(1-\pi) \right) \] \text{ log-log }

2. Poisson \( \mu \)
   \[ y = 0, 1, 2, \ldots \]
   \[ y! \]
   Natural link \( g(\mu) = \log / \mu \)
   \[ g(\mu) = \mu \] \text{ identity }

3. Normal \( \mu, \sigma^2 \)
   \[ \mu = \eta \]
   \[ \sigma^2 = \phi \]
   \[ N = x' \beta \]
   Gauss-Markov + normality

Dispersion parameter \( \phi = \sigma^2 \)

expresses uncertainty
\begin{equation*}
\Gamma(x, \nu) = \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\mu} \right)^x \exp\left( -\frac{x}{\mu} \right)
\end{equation*}

\begin{align*}
\text{E}_Y &= \mu \\
\text{var}_Y &= \mu^2 / \nu \\
\beta &= 1 / \nu \\
\text{Natural link } g(\mu) &= Y / \mu = \nu \\
\text{E}_Y &= 1 / \beta \quad \text{need } \beta > 0 \\
\text{Sometimes use } \\
g(\mu) &= 1 / \mu = \nu
\end{align*}
Inverse Gaussian

\[ f(y; \mu, \sigma^2) = \frac{e^{-(y - \mu)^2 / 2\mu^2} \phi(y)}{\sqrt{2\pi\mu^2} \sigma^2} \quad y > 0 \]

Distributions

\[ EY = \mu \]

\[ \text{Var}Y = \mu^3 \sigma^2 \]

Natural link \[ g(\mu) = 1/\mu^2 = \pi \]

\[ EY = 1 / (x \beta) \]

\[ \gamma = \sigma^2 \]
Selecting and Checking Models.

Goodness of fit statistic.

Peanon statistic,

\[ X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_i^2} \]

\( \hat{\mu}_i \approx \mu \) \quad \text{p = no. of estimated means}

Deviance or likelihood ratio statistic,

\[ D = -2 \ln \sum \left( l_i(\hat{\mu}_i) - l_i(y) \right)^2 \]

\( \approx \sum \chi^2_{n-p} \)

Problems with approx if \( y_i \) is, \( y_i = 0 \)
Selecting Checkpoint Models

Residuals

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

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

l_i(y_i) = \left[ y_i \theta_i - b(\theta_i) \right] \sqrt{\phi} + c(y_i, \phi)

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

Deviance residual

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

Pearson residual

r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i)}}

After highly skewed
need adjustments: A remed...
Variable selection.

Sometimes if many potential covariates

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

Likelihood

\[ l(\beta; \theta) = \prod_{i=1}^{3} l_i(\xi_i; \theta) \]

where \( \theta \) contains other parameters, e.g., \( \beta \)

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

All subsets selection:

Akaike Information Criterion

\[ AIC = -2 l(\tilde{\beta}, 0, \tilde{\theta}) + 2(n + s) \]

\[ n = \text{dim}(\tilde{\beta}) \]

\[ s = \text{dim}(\tilde{\theta}) \]

\( \tilde{\beta}, \tilde{\theta} \) mle under \( H_0 \)
Stepwise backward.

Stepwise forward.
Goodness of fit

**Deviance:** \( D(y; \hat{\mu}) \)

Normal: \( \sum (y - \hat{\mu})^2 \)

Poisson: \( 2 \sum [y \log \frac{y}{\hat{\mu}} - (y - \hat{\mu})] \)

Binomial: \( 2 \sum [y \log \frac{y}{\hat{\mu}} + (n - y) \log \frac{n - y}{n - \hat{\mu}}] \)

Gamma: \( 2 \sum \left[ -\log \frac{y}{\hat{\mu}} + \left( \frac{y}{\hat{\mu}} \right) \right] \)

Inverse Gaussian: \( \sum \frac{(y - \hat{\mu})^2}{\hat{\mu}^2 y} \)

Scaled deviance: \( D^*(y; \hat{\mu}) = D(y; \hat{\mu})/\hat{\mu} \)

**Analysis of deviance**

ANOVA

\[ \frac{D_0 - D_1}{\hat{\sigma}^2 (p - q)} \sim F_{p-q, n-p}, \quad \text{i.e. Gaussian result} \]

May need a better approximation.
Contingency table data,

E.g. lung cancer data

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

E.g. lizard data

Question: Categories independent?
## Table I

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

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

Smoking effect†

|                 | 1.0 | 3.39 | 8.16 | 10.1 | 18.2 | 22.6 | 36.8 |

* Age fit = exp(\(\mu + \alpha_j\)), where \(\mu\) and \(\alpha\) are ML estimates defined by the product model.

† Smoking effect = exp(\(\delta_k\)), where \(\delta\) 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(\(\mu + \alpha_j + \delta_k\)).
### Table 1-2

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

(a) Observed values

<table>
<thead>
<tr>
<th>Perch Height (feet)</th>
<th>Perch Diameter (inches)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 4.0</td>
<td>&gt; 4.0</td>
</tr>
<tr>
<td>&gt; 4.75</td>
<td>32</td>
<td>11</td>
</tr>
<tr>
<td>≤ 4.75</td>
<td>86</td>
<td>35</td>
</tr>
<tr>
<td>Totals</td>
<td>118</td>
<td>46</td>
</tr>
</tbody>
</table>
There are 3 distributions commonly used:

1. Poisson

2. Multinomial

3. Product multinomial

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